[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Ian McMaster

TITLE OF THESIS: A PROPOSAL FOR COMPUTER ACQUISITION OF NATURAL LANGUAGE

DEGREE FOR WHICH THIS THESIS WAS PRESENTED: M.Sc.

YEAR THIS DEGREE GRANTED: 1975

PERMANENT

DATED ..April 27,.... 1975

THE UNIVERSITY OF ALBERTA

A

Proposal

for

Computer Acquisition

of

Natural Language

by

© Ian McMaster

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

IN

Computing Science

DEPARTMENT OF COMPUTING SCIENCE

EDMONTON, ALBERTA

SPRING, 1975

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled A PROPOSAL FOR COMPUTER ACQUISITION OF NATURAL LANGUAGE submitted by Ian McMaster in partial fulfilment of the requirements for the degree of Master of Science in Computing Science.

Date 25 April 1975

To Susy

ABSTRACT

The relationship between linguistics and artificial intelligence should be one of mutual exchange of hypotheses, data, and models. Linguists have provided data and some tentative explanations of some aspects of language acquisition, but no comprehensive paradigms. Several models, including an attempt at a fairly comprehensive one, are reviewed.

The compromise between modelling and pragmatic considerations is examined with reference to the characteristics of a possible language acquisition program. Four existing language acquisition programs are reviewed.

A comprehensive language acquisition program is proposed, its components and strategies are described, and possible implementation methods, some already in existence, are offered.

An experiment with a pilot program for vocabulary acquisition is described. The program uses an artificial environment and natural language input to build associations between words and concepts.

ACKNOWLEDGMENTS

I thank Jeff Sampson for his patience and for the working relationship which he established, without which I would have had difficulty in following creative avenues of research. Gary Prideaux, Kelly Wilson, Len Schubert, Nick Cercone and John Gray have supplied fruitful criticism and ideas, for which I am grateful.

Rici Liknaitzky eased the pain of programming by implementing and hacking MACLISP.

My deepest gratitude goes to my wife, Susan, for her patience and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

1

Introduction

1.1 Language acquisition

Perhaps the most fascinating of the attributes which have separated man from the other species is his possession of language. However, even more interesting is the ability of almost every child to learn a highly developed mother tongue. His accomplishment is more impressive than that of the student who learns a language in the classroom, since the child starts with no linguistic experience and, in general, has no one to tell him anything about the target language.

What the child *is* supplied with is a rich collection of experiences and concomitant verbal intercourse. From these, he learns to understand intonation and words, then uses sounds (which he has been exploring during the babbling stage) to produce words of his own. Between eighteen and twenty months, he goes on to put these words together, first

into pairs, then into longer utterances. From this point on, the child continues to increase his linguistic abilities, not by making crude and mistaken attempts at adult language, but by continually modifying a well-defined language of his own to bring it closer to the one around him.

Language acquisition by children is thus a complex, important, and highly interesting process. It has earned increasing attention from psychologists, linguists, and educators in the past twenty years, and substantial gains have been made in understanding the course of language development. However, present theories of acquisition are models which attempt to explain only aspects of the language learning process. There is no theory of primary language acquisition that draws together these various aspects, diverse schools of thought, and differing sources of data <Dale 1972>.

1.2 Why computer acquisition?

In 1959, a "conversation machine" <Green, Berkeley, Gotlieb 1959> was written as an attempt to understand the problem of satisfying Turing's test for an intelligent machine <Turing 1950>. Since then, computers have been used for a variety of experiments in natural language processing for a variety of reasons. The predominant type of system which has been written can be roughly described as the

"question-answerer". Its long-range goals are:

(a) to understand input text in the language that the user normally uses for written communication;

(b) to process the meaning of the input in order to remember it, carry out actions, or retrieve information with an effectiveness that is the same as, or better than, that of a human being; and

(c) to express any response to the user's input in as understandable a form as possible.

Some reasons for such systems are:

(a) to arrive at a theory of cognition;

(b) to arrive at a theory of language understanding; and

(c) to enable a computer to act as an intelligent and hence useful partner in some practical human endeavour, using normal human means of communication.

It is fairly obvious that if (c) is to be accomplished, (a) and (b) must be at least implicitly attained. Further, a theory of human language understanding must include a theory of acquisition, and the latter is <u>also</u> necessary for the attainment of full mechanical language understanding.

> "One serious criticism that can be leveled at the parsers [of recent natural language systems]...is that they do not learn how to parse; rather they are programmed to parse. Practically, this means that these efforts will never be able to handle

> more than a small subset of any real language like English...What is necessary is a system with basic acquisition devices so that it could program its own parser, given some experience with the language." <Anderson, Bower 1973, page 131>

Just as question-answerers may be tools for finding theories of language understanding, so a computer program which acquires a natural language and perhaps models the child's progress will help to uncover a theory of language acquisition.

1.3 Is it new?

None of the question-answering systems alluded to above has incorporated a language acquisition component. Some have a feature which allows the addition of new words by precise definition of their meaning in either English or some formal language. Important though this is, it contributes very little to language acquisition, since the major problem is not definition in terms of existing words. The core of language acquisition is the formation of new procedures for analysis and generation of language without explicitly being told those procedures.

There are four important systems which have attacked the problem of acquisition. As we shall see in Chapter 3, however, they all suffer from defects that make them poor paradigms for language acquiring systems.

1.3 Is it new?

1.4 Preview

Chapter 2 will look at the relationship between linguistics and artificial intelligence and will give a sketch of the course of acquisition in children as seen by linguists. A selection of explicit theoretical models will be presented and criticized.

The first part of Chapter 3 will attempt to place the possible characteristics of language acquisition on a scale ranging from faithful modelling of the child to practical mechanical learning of natural language. This will be followed by a review of four major computer-implemented natural language acquisition systems, with reference, where possible, to the criteria outlined in the first part of the chapter.

Chapter 4 will present a detailed description of a possible comprehensive language acquisition program. Implementation of each component of the system will be examined and some solutions will be proposed where none exist or existing ones are inadequate. The design of such a system will implicitly contain linguistic-theoretic assertions, and I will comment on their place in linguistics.

A pilot program of part of one of the strategies in a complete language acquisition program will be described in Chapter 5. VAS, a vocabulary acquisition system, has been

written and used in an experiment concerning the acquisition of the meaning of words. I will describe the way in which the environment is modelled and the way knowledge of the world and the system itself is stored. The system is given the capacity to focus on parts of its environment as it is receiving linguistic input, and it uses these focus-utterance pairs to construct its vocabulary. The results of the experiment will be reported, and the place of VAS in the parse-evaluate-modify loop of the complete system will be proposed. The chapter concludes with a discussion of further possible research involving VAS.

Chapter 6 will summarize the purpose of the work and draw some general and critical conclusions about its success and potential for success.

2

The Linguist's View

2.1 Linguistics and Artificial Intelligence

The relationship between the fields of linguistics and artificial intelligence (AI) should be a reciprocal one. As a result of observations of human language behaviour, a hypothesis is made about the relationships between utterances, and claims about a model of the acquisition process are associated with the hypothesis. These claims, if they are formalized properly, can be incorporated in a computer program. The linguist also provides a description of the input, i.e. the stimuli, linguistic and otherwise, to which the human is exposed. By providing the program with input which is related in a formal way to that of the human, and observing the performance of the program, the hypothesis can be tested and modified by the linguist in light of the results <Kelley 1967; Schwarcz 1967; Wheatley 1970>.

Consider the following example of this process. Bloom <1970> has pointed out that it is impossible to observe

directly the relationship between the semantic structures of the child and his language. If artificial intelligence is to be considered a serious goal, we must assume that we can construct a formal system capable of representing perceived reality on some level, in a way which is independent of natural language. Hence we can in principle observe the effect of the semantic representations in this formal system on language performance.

As another example of the linguistic utility of a computer model, consider the search for substantive linguistic universals. A machine model of the human language user offers an opportunity unparallelled in the human subject: the same machine model could acquire two different languages each as a first language, using the same initial knowledge and receiving the same non-linguistic input.

If a linguistic model proves effective in reproducing language behaviour in a computer program, it may be used to pursue the goal of a useful language understanding system.

Information may flow in the other direction. The AI researcher uses heuristic techniques in the programs he writes to comprehend and produce natural language. The linguist may then infer hypotheses from these techniques and test them for compatibility with his observations.

The ideal described above is, of course, far from being

realized. Consider language acquisition as an example. Observations of the environment and utterances of children are far from comprehensive. Much of the data comes from longitudinal studies, often of offspring of the researchers themselves <Kelley 1967>. There are inherent problems in observing children. It is difficult to know at what level to describe the environmental context of utterances. Since cognitive development is not precisely understood, it is difficult to infer what the non-linguistic input is; semantic origins of the child's utterances are similarly obscured. Furthermore, it is difficult to characterize precisely the idiolect of a child at an instant in time. The idiolect must be inferred from a number of utterances, and these utterances must necessarily be observed over a period of time during which the idiolect itself may be changing. There are scores of other obstacles in the way of a comprehensive account of linguistic phenomena <Kelley 1967>.

From the point of view of AI, there are two major defects in models proposed by linguists. The first fault is exemplified by the statement that

> "When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way, to construct such a derivation." <Chomsky 1965, page 9>

The so-called competence model described by transformational-generative grammar (TGG) is thus purely a

formal system for organizing natural language utterances. If one attempts to embed this "underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance" <Chomsky 1965, page 4> in a performance model, one is faced with two problems. First, there is no *a priori* reason why it should be possible to construct a performance model with a TGG as an integrated, usable component. Second, Chomsky's theory does not specify in any useful way what the characteristics of the performance model would be which would relate a TGG to a general control structure or the equally vital analytical component <Derwing 1973>. Add to these lacunae the problem of generating sentences which are appropriate to the situational context, and the competence *model* is relegated to the status of an ingenious, neat notational device of dubious utility to AI.

Vagueness is the second fault of linguistic models, and it is more characteristic of hypotheses from rationalists than from empiricists. As we shall see in section 2.3, rationalist models of performance (in particular acquisition models) are not usually precisely formalized and quantified, and lack specification of details which are vital to implementation and testing.

The reason for the imprecision of rationalist models

may be the absence of a unifying paradigm[1]. There has been relatively little detailed speculation on a general formal theory which might govern the relationships among all the major components of the individual and his experience: linguistic and non-linguistic input, non-linguistic cognitive development, semantic representations, and analytical and generative mechanisms.

The next stage in the interchange, namely incorporation of a linguistic hypothesis in a program, has become less difficult over the last ten years. As we shall see in section 4.4, there have appeared a number of systems comprehensive enough to supply a framework within which fragmentary hypotheses could be tested. The major exception is the phonological component. Because of this deficiency and for other reasons to be given later, deciding on the nature of the input to the system is as serious a problem as that of finding an internal structure into which to incorporate a hypothesis; it is a problem which AI models in general must face, and sections 3.1.1 and 3.1.3 examine it in detail.

Interdisciplinary communication breaks down most in the flow of results from AI to linguistics. Some linguists deny the relevance of AI to their research:

"...the perceptron models, heuristic methods, and

---

[1] "Clearly, there is no well-defined paradigm for the study of language acquisition." <Moore 1973, page 5>

> 'general problem-solvers' of the early enthusiasts of 'artificial intelligence' are successively rejected on empirical grounds when they are made precise and on grounds of vacuity when they are left vague..." <Chomsky 1968, page 79>

> "LAD[Language Acquisition Device], of course, is a convenient fiction. The purpose in considering it is not to build an actual machine." <McNeill 1971, page 20>

Most linguists simply ignore AI or, like Dale <1972>, give it perfunctory mention, with some exceptions:

> "The writing of grammars of children's language can only tell us that a change has occurred. Such grammars cannot tell what the changes have been in how the child processes sentences, nor how the changes have come about. For these reasons, new studies are likely to develop models which can be tested. Language processing structures and language acquisition systems are going to be where the action is." <Ervin-Tripp 1971, page 212>

2.2 A sketch of acquisition

Lacking an adequate formal theory of language performance[1], it is impossible to give an adequate formal definition of language acquisition. About all that can be generally agreed on is that when a child is born he cannot communicate his thoughts using the linguistic means of the adults around him, and that he has acquired language when he can "communicate successfully on a variety of topics of discourse with other members of his 'linguistic community'"

---

[1] "...the goal of a comprehensive and rigorous theory of linguistic performance is one that is not about to be achieved for a long, long time." <Schwarcz 1967, page 39>

<Schwarcz 1967, page 41>.

There are two independent ways in which this process can be broken into subprocesses. First, acquisition of the language spoken around a child can be viewed as acquisition of a sequence of idiolects, the last of which is "closer" by some metric to the target language than the first. Second, understanding and production can be separated, since the class of utterances understood at any point is, in general, different from the class of utterances produced by the acquirer <Kelley 1967; Lenneberg 1969; Slobin 1971b; Dale 1972; Schlesinger 1971>. In particular, it seems likely that children understand some parts of utterances (that is, they extract at least some of the intended meaning) before they reach the stage of producing their own first word.

To give an account of acquisition, it is necessary to use an observationally adequate system to describe the idiolects of the child. This prerequisite is agreed on by both rationalists and empiricists <Chomsky 1968; Staats 1971; Derwing 1973>. Linguists have provided four major theories of language which could be used to describe the development of language:

(1) transformational-generative grammar (TGG) <Chomsky 1965>,

(2) generative semantics <Lakoff 1971>,

(3) case grammar <Fillmore 1969>, and

(4) systemic grammar <Turner, Mohan 1970; Winograd

1971; Halliday 1973>.

Of these, (1) is by far the most well-developed, (2) and (3) are more recent descendants of Chomsky's transformational theory, and (4) is used by few linguists outside Europe.

Most linguistic research in language acquisition has been within the Chomskian TGG framework, with some notable exceptions. Some recent work <Bowerman 1973> has used case grammar, and Brown <1973> has claimed that a rich (that is, explicitly semantic) description is most appropriate for the acquisition period. Bowerman's description is of Stage I (1<MLU[1]≤2), and I shall follow her case grammar description for this stage. Prior development has not, of course, been given grammatical treatment, and subsequent stages have almost exclusively been treated in terms of Chomskian TGG; I shall do so too.

During the first year, the child babbles, producing, among others, sounds he will eventually use when he begins to speak. Since phonology is relatively unimportant to the system to be proposed, an analysis of this aspect of development is not required here. Accounts and theories of such development are reviewed by Dale <1972>.

There is evidence that during the stage before speech occurs there is understanding by the child of some adult speech <Dale 1972>. Between ten and twelve months, in most

---

[1] Mean Length of Utterance

cases, the child for the first time utters a meaningful word. That is, he uses the word consistently, spontaneously (i.e. not imitatively), and in appropriate situations. These utterances have been called "holophrastic" because they are apparently intended to convey meanings which would normally, in adult speech, be conveyed by a sentence. In general, however, linguists have refused to consider one-word utterances as language, for the reason that such utterances can be given no structural description beyond the trivial one:

S
|
<word>

If we restrict ourselves to a grammatical description, this position may be justifiable, but an acquisition model, being one of performance, must be concerned with how the child modifies his production methods to produce utterances combining words which he has previously used only singly, and so must attempt to explain holophrasis. It is plausible that the child's cognitive development is an important factor at this stage, and that it is the interpretation of conceptual structures by the language component that constrains the form of utterances. I will expand on this hypothesis in Chapters 3 and 4.

Between eighteen and twenty months, though this is highly variable, the child's MLU rises above 1. Roger Brown has named the interval between this point and MLU=2 "Stage I", and has characterized it as the period during which the

child starts to represent semantic roles and grammatical relations by grammatical relations in his utterances <Brown 1973>. There is not, of course, a sharp boundary between this and the next stage, and there are some utterances near the end of Stage I which explicitly contain the modulations, such as tense, aspect, and number, which are characteristic of the next stage.

Bowerman <1973> has used a case grammar to describe the cross-linguistic characteristics of sample American, Finnish, Samoan, and Luo children. In this description the symbols stand for terms as follows:

S	sentence
M	modality
Q	interrogation
Neg	negation
P	proposition
V	verb
A	agentive
O	objective
L	locative
D	dative
E	essive

Fillmore <1969> and Bowerman <1973> have described the meaning of these terms. Stage I deep structures are described by the grammar in figure 2.1. In late stage I, three-word utterances and the factive case begin to appear.

Although lexicons are obviously idiosyncratic and language- and culture-specific, features such as <±animate>, <±pronoun>, <±directional>, <±quantifier> occur in almost all Stage I grammars. Ordering transformations are inherent in case grammar descriptions, since word order is the

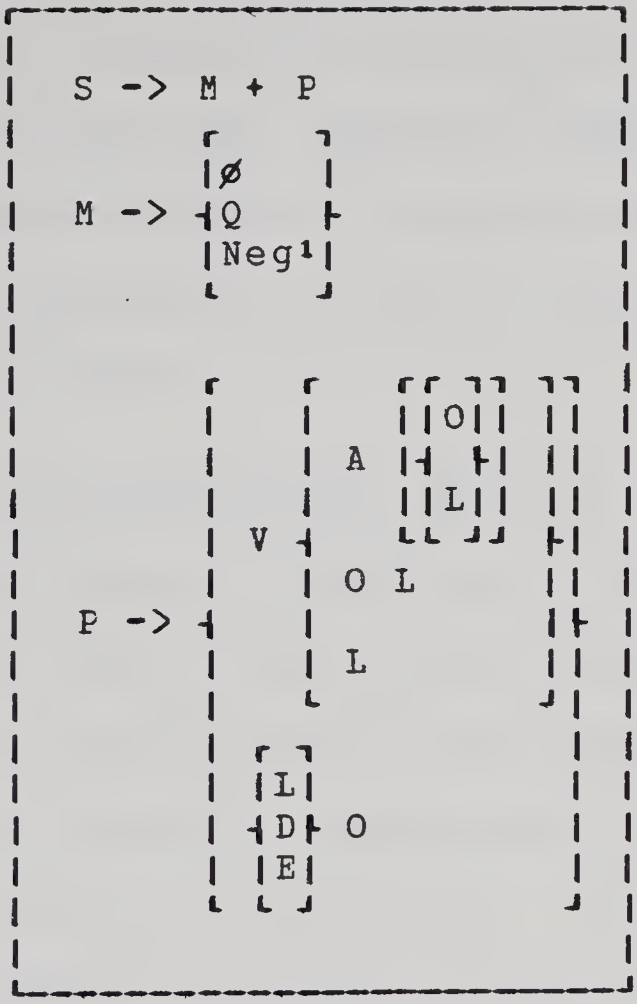


Figure 2.1 Phrase-structure component of a cross-linguistic case grammar of Stage I speech <Bowerman 1973, page 210>
[1]for some children in very simple utterances

predominant way in which children in Stage I represent semantic roles <Brown 1973>, and case grammar deep structures are unordered. Since Fillmore's formulation requires a verb in the deep structure <Fillmore 1968>, a verb-deletion transformation is required during this stage. However, the lack of a verb in many of the sample utterances probably points more to deficiencies in Fillmore's theory than to the existence of such transformations in the child's production system <Brown 1973>. Despite the fact that case

grammar theory has not been developed to nearly the same extent as Chomsky's theory, its attraction lies in the way it seems to capture the most important phenomenon of Stage I: the direct correspondence of surface constituents and their order with the semantic roles of referents in the environment of the child.

Little detailed naturalistic data has been gathered from children past Stage I. Based on samples from three American children, Brown <1973> has characterized Stage II (2<MLU≤2.25) as the period during which grammatical morphemes that mark semantic modulations of the simple sentence begin to appear productively. Unfortunately, he is unable to give a precise specification of what defines these grammatical morphemes. He admits that "they may not constitute a single class semantically" <Brown 1973, page 254>. Figure 2.2 shows the 14 morpheme classes (which Brown calls morphemes) considered in Brown's study of American children. Many of these appear at the same stage in Luo and Russian. Brown did not analyze data from other languages because the techniques for collecting, organizing, and judging it do not allow comparison with his. However, other data indicate that the order of development of his "fourteen morphemes" is predictable for children acquiring Standard American English. The present progressive and the prepositions "in" and "on" were observed to be under control by Brown's subjects by the end of Stage II.

Morpheme	Average rank
1. Present progressive	2.33
2-3. _in_, _on_	2.50
4. Plural	3.00
5. Past irregular	6.00
6. Possessive	6.33
7. Uncontractible copula	6.50
8. Articles	7.00
9. Past regular	9.00
10. Third person regular	9.66
11. Third person irregular	10.83
12. Uncontractible auxiliary	11.66
13. Contractible copula	12.66
14. Contractible auxiliary	14.00

Figure 2.2 Mean order of acquisition of 14 English morphemes across three children <Brown 1973, page 274>

The only published grammatical descriptions of Stage II speech which I have found are TGG's. Brown has extended the Jacobs-Rosenbaum <1968> formulation of TGG and sketched a grammar of the fourteen morphemes. The deep structure rules are:

$$S \rightarrow NP\ Aux\ VP$$

$$NP \rightarrow \left\{ \begin{array}{l} [Art]\ N\ [S] \\ NP\ S \end{array} \right\}$$

$$VP \rightarrow VB\ [NP] \left[\left\{ \begin{array}{l} NP \\ S \end{array} \right\} \right]$$

Control of the morphemes comes about as lexical features, segment structure features, and transformations are incorporated into the grammar. For instance, the plural

morpheme requires that nouns be marked <±common>, <±count> in the lexicon. A segment in the deep structure may contain the feature <±singular>, through the application of a rule such as:

N -> <+N>
<±singular>.

The plural suffix is introduced by the Noun Suffix transformation, which is blocked if the noun is one which has an irregular plural realization, like *woman*, *man*, or *child*. The Determiner Agreement transformation copies the number feature of the noun to the determiner segment. The Predicate Nominal Agreement transformation has to copy the number from the subject to the predicate if the verb is a copula, and there must also be a device to ensure that the predicate nominal segment does not acquire number in some independent way.

The above transformations introduce features into the underlying structure of a sentence. The Noun Suffix transformation includes the feature <+affix> in the suffix segment, and the Determiner Agreement transformation includes <+article> in the article segment.

The above-described mechanisms are typical of the fourteen morphemes which appear in Stage II.

There are some interesting phenomena in the acquisition of English at this stage that may be relevant to a machine model. These are the precedence of uncontractible over

contractible forms, errors in the use of articles, and overregularization. In Brown's sample, the uncontractible forms of the copula and auxiliary like _is_, _am_, and _be_ in

    Here I am
    There it is
    I be quiet

were controlled long before the contractible ones, which were omitted for a long time. The slower development of contractible morphemes may be related to their low acoustic perceptibility, and is also predicted by conventional grammatical description which derives contracted forms using an extra transformation, increasing their derivational complexity.

Brown points out an interesting pattern of error in the use of the definite and indefinite article. The wrong choice of article is made most often in the case where the referent is specific for the child and nonspecific for the listener; that is, where the article should be "a". The error may be symptomatic of the egocentric stage, in which the child cannot, or does not, see the world through another's eyes. This in turn may be an instance of a more general phenomenon of importance to machine models: the dependence of children's acquisition on conceptual and emotional development.

Irregular past tenses are in general controlled before the regular ones <Dale 1972>. However, when the regular past inflection is learned there is a short period during

which the child inflects with a regular suffix verbs which require an irregular past inflection. There are a number of implications for a machine model. For instance, rules which are discovered to be powerful, like the regular past inflection, may be hypothesized for all "similar" cases. Or it may be that rules which require blocking in special cases are always developed without such blocking at first. Or it may be that features which are used to give a morpheme a peculiar surface form are of a type which is different from that of features of a more semantic (and general) sort, like <±count>, <±animate>, *et cetera*; hence they are treated in a different way (perhaps even overlooked) in developing new rules.

Brown's major conclusion about the determinants of acquisition is that it is semantic and grammatical complexity, and not frequency of tokens, which directs the course of development of the grammatical morphemes characterizing Stage II.

Stage III is characterized by the appearance of modalities of simple sentences. That is, interrogatives, negatives, and imperatives appear, first with the simple expression of the Q morpheme by a word (in most TGG descriptions of English usually in the initial position). While yes/no questions are generally answered appropriately by the child, wh- questions are not <Dale 1972>.

Later, Auxiliary Shift and Do-formation transformations

appear in yes/no questions, but sometimes they are applied incompletely, as in

Did you broke that part?

Wh- questions develop first without Auxiliary Shift, to produce

What you have in you mouth?
Why you smiling?

Note that the auxiliary is missing in the why-question, whereas it should be under control at the end of Stage II, according to Brown <1973>. However, a single instance of error is not anomalous, since Brown's definition of control is that the morpheme is produced correctly in at least 90% of all contexts in which it is clearly required, in 6 consecutive hours of speech sampling. Sarah, one of Brown's subjects, still erred in her use of the auxiliary in Stage V.

Throughout Stage III, the child responds appropriately to questions which are more complex than those he is uttering <Dale 1972>, a phenomenon which lends support to the separation of comprehension and production in an acquisition model.

Negation is another modality which begins to appear productively in Stage III. Bloom <1970> has subclassified negation as expressing variously nonexistence, rejection, and denial, and it is in this order that linguistic control of these concepts is acquired in English. As with questions, negation in English starts with a simple overt

negative morpheme in the initial position, as in

No go in
No more apple.

Transformations which insert "do" and move Neg to the auxiliary follow.

The interesting thing is that denial was expressed using simpler syntactic means than the other two senses of negation, but developed last. Furthermore, as each sense appeared, it was expressed using the most primitive syntactic means, even though more complex rules had already been developed for other negatives. At Time I, Kathryn <Bloom 1970> expressed nonexistence often, exclusively by a sentence-initial "no", as in

No pocket
No turn

and there was only one expression of denial, namely

No dirty

At Time II, nonexistence was expressed with sentences as complex as

Kathryn no fix this
Man no go in there

whereas denial, which was expressed with significant frequency, was always expressed by the primitive form:

No ready

At Time III, the frequency of denial increased by 200%, and its form advanced in complexity to

I not tired
That not a blue one.

The development of negation may again point to the powerful effect of conceptual development on the course of linguistic development. Although a theory of semantic complexity is lacking, there are cogent reasons for believing that denial is a more difficult concept to process than the other two <Brown 1973>. It requires the handling of two propositions at once, one which has just been affirmed and which the child is denying, and the other the affirmation which the child is making about his view of the world. McNeill and McNeill <1968> have called this relationship "entailment-non-entailment". The fact that English-learning children use the most primitive syntactic means each time they acquire a new sense of negation may indicate that grammatical rules depend for their invocation on the semantic origin of the utterance.

Detailed observation and analysis of development beyond Stage III is difficult to find in the linguistic literature. Brown <1973> implies that the next two stages are characterized by the introduction of sentence embedding and sentence coordination, in that order. He claims that among the first embedded sentences to appear are object noun phrase complements, as in

I hope I don't hurt it

embedded wh-questions, as in

Know where my games are?

and relative clauses, as in

That a box _that they put it in_.

Limber <1973> has observed the same phenomenon in three American children, and offers two further general hypotheses. The first is that if a child acquiring English has reached the four-word utterance stage, and if a new complement-taking verb is learned, then within a month the child will produce a complement clause with that verb, as in

Watch me _draw circles_.

The second hypothesis is that such complement-taking verbs appear in a characteristic order, namely:

(1) "want" and "watch" groups

(2) auxiliaries

(3) verbs taking wh-clause objects, like "show"

(4) verbs taking propositional objects, like "think".

Limber also hypothesizes an order of appearance of relatives, in terms of the nouns to which they are attached:

(1) abstract adverbial nouns, like "place" and "way"

(2) empty noun heads, like "thing", "one", and "kind"

(3) common nouns like "ball" and "cheese".

In addition, the relative conjunction is first ø, then "that".

Coordination can be applied to full sentences, as in

You snap and _he comes_

cr to two or more sentences which are partially identical, partially different, as in

Mary sang and danced.

It should be clear from this sketch that the early stages of acquisition are the most studied and best-understood ones, but that even in these there are many questions left unanswered with respect to both the grammatical description and the determinants and psychological processes of acquisition.

2.3 Linguistic models

The broad questions that a theory of human language acquisition must answer are:

(a) with what psychological abilities and linguistic propensities does the child start?

(b) what is the relationship between the development of general motor skills, perception, and concepts, and the development of language skills?

(c) what is the relationship between the non-linguistic and linguistic environments and how does it induce comprehension?

(d) what is the relationship between ccmprehension and production?

Answers to question (a) come from two schools <Katz

1966>. Empiricists claim that the building blocks from which concepts are formed are simple weighted associative bonds, and that there is hence no limit on the character of ideas or concepts which can be learned by humans. Language is acquired by application of these general techniques to linguistic experience, and there are no special mechanisms for gaining linguistic competence. The rationalist or nativist view is that there is innate specification of psychological structures necessary for primary language learning, and that acquisition is the process by which linguistic exposure and experience turns this innate capacity into linguistic competence. Arguments for the two positions are presented by Church <1961>, Katz <1966>, Ervin-Tripp <1971>, Staats <1971>, and Dale <1972>.

The AI researcher must choose strategies from these two positions on pragmatic grounds. There is no adequate general theory of learning, and further, a successful general learning program has yet to be written. Hence, the choice is somewhere between the extremes of trying to write a general learning program and trying to write a program specifically for language acquisition. If the former attempt produced a system which successfully learned language in the same way as other skills, support would be lent to the empiricist position to the extent that any observationally adequate model lends support to a psychological theory. Success of the latter attempt would support neither position over the other.

The rationalist linguists may also be separated on the basis of their methods of description of language. Among these are: Chomskian transformational-generative grammar; generative semantics; case grammar; and systemic grammar. A systemic grammar for English children of age 5 has been written <Turner, Mohan 1970> and, as mentioned in section 2.2, Bowerman <1973> has written a cross-linguistic case-grammar description of Stage I speech, but by far the bulk of rationalist theories and studies of acquisition have been in the context of Chomskian or neo-Chomskian TGG.

Most language-processing models in AI are implicitly based on rationalist models, and the paradigm to be presented in Chapter 4 is no exception. For this reason, models mentioned below are rationalist. My methodological position will not be presented in the following critiques, except as implied by my evaluation of the characteristics of the models. Detailed proposals for a paradigm for a language acquisition system will be presented in Chapter 4 and to a lesser extent in Chapter 3. While there are other models of various fragmentary aspects of acquisition, the following constitute a representative sample.

2.3.1 The Chomsky school

The transformational-generative theory of language as presented by Chomsky <1957> and later extensively elaborated

by him <Chomsky 1964; 1965; 1968; 1970> has provided a tool of enormous importance to accounts and theories of language acquisition. It has allowed linguists to set forth the phenomena which occur during the development of child language in a precise and unified way <Dale 1972; Bowerman 1973; Brown 1973>. However Chomsky <1968> claims a higher status for transformational theory <Derwing 1973>. He postulates first that a universal theory of grammars, that is, a theory which describes the possible form and characteristics which any generative grammar for a human language could have, is also a statement about the the innate characteristics with which the human mind approaches language acquisition. His second claim is that a particular generative grammar is a model which specifies a structure which has been built within the mind of the speaker of the corresponding language. The latter claim disengages him from the behaviourist school; the former labels him as a strict rationalist.

Chomsky may indeed be correct in his argument that the empiricist view turns out to be untenable whenever it is pinned down and vacuous where it is vague. For the researcher in artificial intelligence, this is not the most important question, for reasons given in section 2.3. What _is_ important is the fact that generative grammar, in particular transformational grammar, is assumed, not chosen, as the model for man's linguistic capacity. Certainly a TGG provides for a language a description which is concise and

unified. However, as an explanatory vehicle, it contributes almost nothing to a theory of comprehension or analysis which is crucial to an explanation of adult linguistic competence, let alone a theory of acquisition <Schwarcz 1967; Anderson, Bower 1973>.

There is also the question of what the process is by which a human goes from a supposed conceptual structure representing intent, information, or whatever the meaning of an utterance is, to the deep structure which is the starting point of the transformational cycle. This question of the semantic origin of utterances is perhaps the most important one that can be asked of an explanatory theory of language, and in recent studies the correspondence of semantic and syntactic structures has emerged as an extremely important phenomenon in early acquisition <Ervin-Tripp 1971; Bowerman 1973; Brown 1973>. Yet the "semantic projection rules" constitute the weakest (in terms of explanation) component of TGG's. This weakness is undoubtedly due to the fact that adequate formal theories of semantics are lacking, but to advance TGG's (with these explanatory inadequacies) to the exclusion of other models seems to smack of the very *a priori* dogmatism which Chomsky <1968> decries in the empiricists.

There are few concrete proposals by Chomsky for a theory of acquisition beyond the metatheoretical. He has stated two criteria:

> "A perceptual model that does not incorporate a descriptively adequate generative grammar cannot be taken seriously. Similarly, the construction of a model of acquisition (whether a model of learning, or a linguistic procedure for discovery of grammars) cannot seriously be undertaken without a clear understanding of the descriptively adequate grammars..." <Chomsky 1964, page 114>.

His comparison of the child to the linguist or grammarian vaguely suggests a "hypothesis-testing" model, in which the child makes hypotheses about phrase-structure rules, features, transformations, etc., and tests them during his linguistic experience, discarding inconsistent hypotheses. There are, however, no suggestions as to procedures or algorithms which might accomplish such a task.

Katz has tried to define somewhat more precisely the principles of Chomsky's model. His definition of the rationalist position is that

> "the language acquisition device contains, as innate structure, each of the principles stated within the theory of language. That is, the language acquisition device contains,
>
> (i) the linguistic universals which define the form of a linguistic description,
> (ii) the form of the phonological, syntactic, and semantic components of a linguistic description,
> (iii) the formal character of the rules in each of these components,
> (iv) the set of universal phonological, syntactic, and semantic constructs out of which particular rules in particular descriptions are formulated,
> (v) a methodology for choosing optimal linguistic descriptions,..." <Katz 1966, page 269>

In his opinion, Chomsky's conception is that

> "the child formulates hypotheses about the rules of the linguistic description of the language

> whose sentences he is hearing, derives predictions from such hypotheses about the linguistic structure of sentences he will hear in the future, checks these predictions against the new sentences he encounters, eliminates those hypotheses that are contrary to the evidence, and evaluates those that are not eliminated by a simplicity principle which selects the simplest as the best hypothesis concerning the rules underlying the sentences he has heard and will hear. This process of hypothesis construction, verification, and evaluation repeats itself until the child matures past the point where the language acquisition device operates." <Katz 1966, page 269>

For the AI researcher this outline is still vague. What is the nature of the "hypotheses" selected from the theoretically unbounded number of possibilities? Katz would answer that they are constrained by the human brain, but that is not a valuable answer unless accompanied by specific suggestions as to the character of such constraints. What among the infinity of possibilities are the predictions made? Does the child check them against <u>all</u> sentences, correct sentences, or parts of sentences, and to what degree is a hypothesis confirmed or infirmed by the evidence? Answers to these questions require more information from linguists to enable AI researchers to make educated choices of mechanisms and algorithms to include in their systems.

2.3.2 <u>Syntactic categories and their relations</u>

Among those who have used Chomsky's conception of acquisition as a starting point for a more precise model, McNeill has a fairly extreme view of innate linguistic

mechanisms. He makes strong claims about the universality of some components of TGG's <McNeill 1971>. First, some syntactic categories such as sentence, noun phrase, and predicate phrase, and the relations which hold among them, are universal. Second, since deep and surface structures are separated in all languages, "...every language is transformational... Although the number of transformational rules in any language is large, the number of universal elementary transformations is a mere handful." From these premises McNeill infers, in a Chomskian way, that these universals - the syntactic categories, grammatical relations, and universal elementary transformations - are present in the pre-linguistic child.

McNeill presents two hypotheses to explain how the child assigns words to syntactic categories. Differentiation results from classifying words into ever more subordinate divisions of the hierarchy of categories. He rejects this hypothesis on the grounds that it does not explain the observed phenomenon that words which eventually end up in one syntactic category, for instance the adjective category, sometimes start in two or more different categories, say the open and pivot categories. Instead, he advances feature-assignment as a mechanism for classification. If the child understands part of an utterance, he makes use of some of the innate relations between syntactic categories. If a processed word is assigned to category A, it is assigned the feature +A. If

it appears after a word of category B, then it is given the feature +B__, and hence has the dictionary entry [+A,+B__]. Thus the phenomenon which is not accounted for by differentiation can be explained by mere removal and addition of features.

The above proposal attempts to explain the formation of the base component of the child's grammar. McNeill goes on to make some tentative remarks about the acquisition of transformations. He hypothesizes that the child acts like a linguist applying simplicity criteria and creates a transformation when phrase-structure rules get tco complex and unwieldy. The exact criteria for introduction of transformations are not suggested, however.

As Ervin-Tripp <1971> has pointed out, McNeill's model is an appropriate one for some of the later non-semantic categories like auxiliary, but it fails to capture the coincidence of semantic relations with grammatical relations that is observed in the child's early language.

2.3.3 Semantic relaticns

Where McNeill has attributed direct analogues of grammatical elements to the child's language acquisition machinery, Schlesinger <1971> has postulated a semantic or conceptual origin for syntactic patterns. His model tries to explain the origin of an utterance as a deep structure

which he calls an input-marker or I-marker (for "input" to the base rules) and which is different in character from Chomskian deep structure.

An I-marker is a nested set of relations; for instance, for

Jchn catches the red ball

the I-marker is roughly

Ag(John,[Ob([Det(the,[Att(red,ball)])],catches)])

This structure is the input to "realization rules" which assign position to the elements of the I-marker and assign them to grammatical categories. For the abcve example, we would obtain the tree in figure 2.3.

The relations in the I-markers are binary. The motivation for this is the explanation of the early production of two-word utterances. Such utterances as "Bambi go", for instance, can be interpreted as direct expression of the relation Ag(Bambi,go). The acquisition of the words themselves is not explained, however.

There two important things to note about the model. First, it tries to mirror performance, not competence, hence there is no neccessary incompatibility with TGG's or any other descriptive theory. Second, and perhaps more important, an attempt is made at specifying a formalism for the meaning of an utterance, and this meaning is seen as the origin of the utterance, not just an adjunct of syntactic deep structure as it is in TGG's. The implication is that

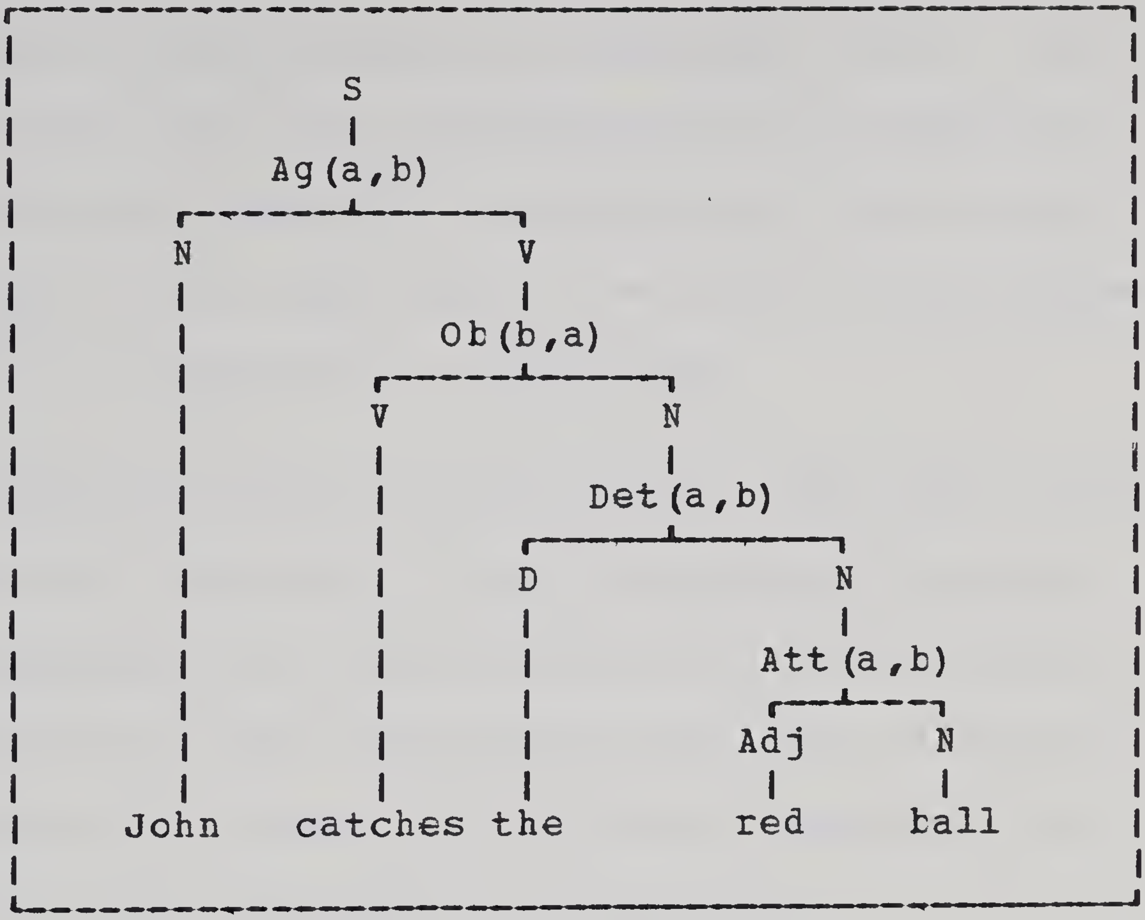


Figure 2.3 The tree representing "John catches the red ball."

I-markers have much in common with more general non-linguistic conceptual structures, and that the realization rule is the linguistic formalism which is innately specified. This contrasts with McNeill's view that syntactic categories, grammatical relations, and elementary transformations are the innate tools.

2.3.4 An information-processing model

While McNeill and Schlesinger make no explicit attack on Chomsky's hypothesis-testing model, and in fact present compatible models, Braine <1971> represents a view which

argues strongly against it. His first objection is that hypothesis-testing requires a non-noisy input. That is, it must be the case that grammatical and ungrammatical utterances are identified correctly for the learner. If not, then he can never hope to be able to use feedback to validate or invalidate a hypothesis.

Braine's second argument is that the child receives no information about what is *not* grammatically correct. Hence if a hypothesis is overinclusive, that is, it generates all acceptable and some unacceptable strings, it cannot in principle be narrowed, since feedback does not tell the child what the unacceptable strings are.

Having argued that in principle the hypothesis-testing model cannot work with noisy input, Braine describes an experiment with an artificial language which, he claims, shows that humans learn language just as well with noisy as with non-noisy input. The conclusion is that humans do not use the hypothesis-testing method.

Braine's argument involves a number of assumptions. One is the implicit denial of any kind of associative mechanism which would strengthen hypotheses as they are used more often. Rules that make the grammar overinclusive would atrophy because the sequences they produce never occur in adult language, making the rules useless for parsing. Kelley <1967> has written a program which uses a hypothesis-testing strategy to learn a grammar of a subset of English,

and it acquires the grammar even with a large number of unmarked ungrammatical input sentences (see section 3.3.2). So Braine's argument that negative information is necessary for a hypothesis-testing model is still very open to question.

The alternative proposed by Braine could be categorized as an information-processing model. As shown in figure 2.4, it consists of a scanner, a series of intermediate stores, and a permanent store. The scanner has the ability to recognize certain pre-defined characteristics of input strings, such as juxtaposition, ordering, and co-occurrence. As these characteristics are perceived in input utterances they are put into the first intermediate store, or move from one intermediate store to the next if they have been seen before. Decay mechanisms in the intermediate stores filter out random or infrequent structures, allowing well-established patterns to reach the permanent store. Once in the permanent store, these structures can then be used by the scanner as templates to match against succeeding input strings.

The above model avoids the previously mentioned shortcomings of the hypothesis-testing model. Noise in the input is filtered by the intermediate stores, and over-inclusive rules are modified by the accretion of more and more specialized patterns. Unfortunately, important details are lacking. The types of patterns recognized by the

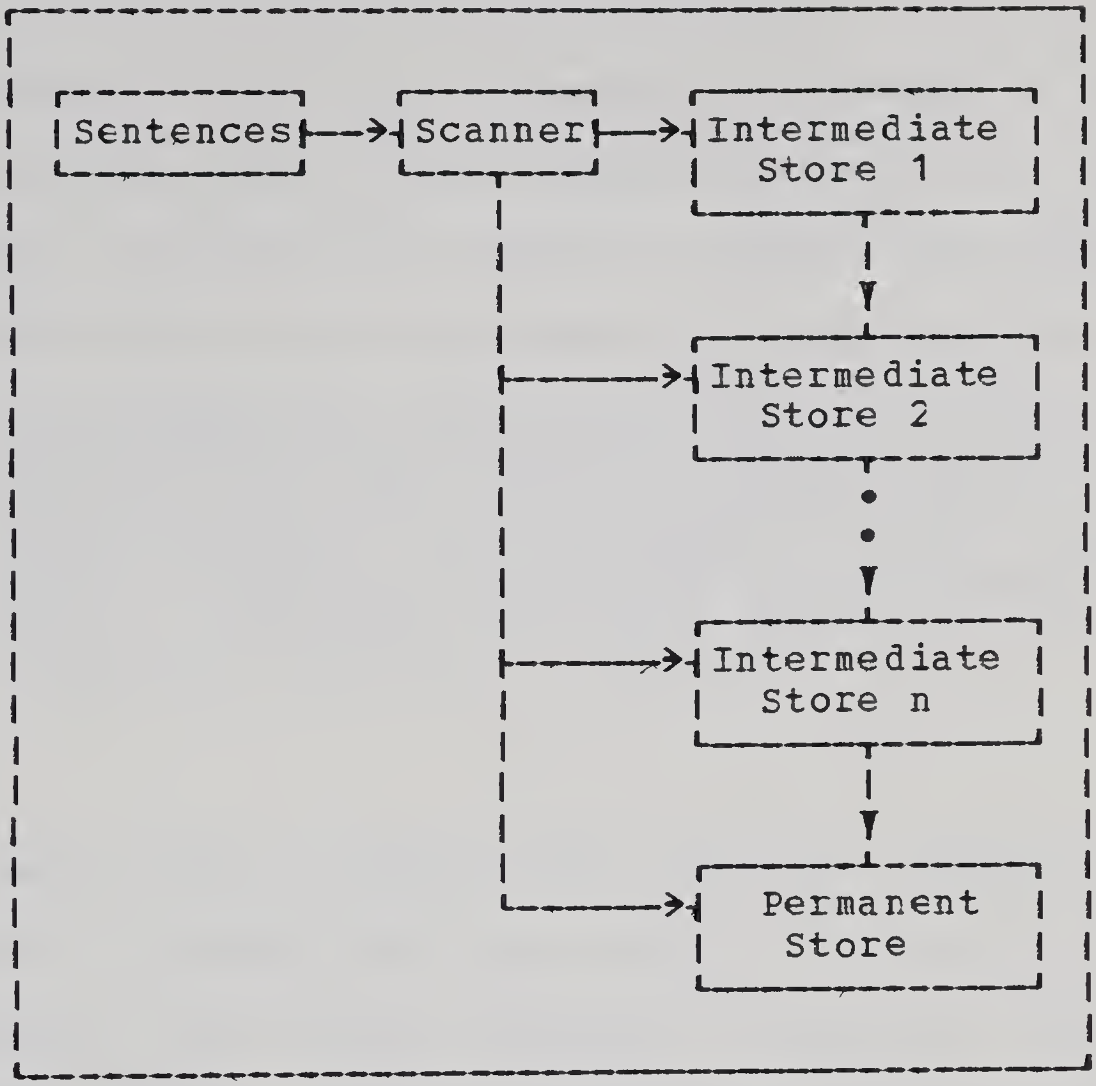


Figure 2.4 Structure of Braine's acquisition model

scanner are not listed, decay characteristics of the intermediate stores are not specified, and nothing is said about changes in these characteristics with time. The relationship of semantics to the scanner is mentioned, but no mechanism for linking them is given. Braine's model is obviously intended as a component in a larger system, since it contains no mechanisms for morphemic analysis or learning, vocabulary growth, or production of utterances, mechanisms which are essential in a comprehensive acquisition theory.

2.3.4 An information-processing model

2.3.5 A comprehensive acquisition model

Schwarcz <1967> has attempted to delineate a computer-implementable model of the "typical speaker" of a natural language, including the ability to acquire the language. Since the model is to represent the typical speaker, it

> "will understand and produce utterances...of a 'representative idiolect' that changes continually over time, and that, after a certain initial training period, will be extensive enough for the model to communicate successfully on a variety of topics of discourse with other members of its 'linguistic community'" <page 41>.

The process of acquisition (Schwarcz' "initial training period") is divided into five stages. The first involves recognition that certain sequences of primitive input elements (sounds, classes of sounds, or symbols) represent lexical items, through "some variety of 'clustering procedure'". The second stage is the establishment of relations between lexical items and individuals, classes, and relations, using extra-linguistic feedback. In the third stage linear orderings of lexical items are associated with relationships in the non-linguistic environment, again requiring extra-linguistic feedback. These associations are generalized in the fourth stage to become functions which map classes of word patterns into their semantic counterparts, using "inductive generalization capabilities built into the model". Schwarcz calls the fifth stage the "transformation learning" phase, in which the model acquires

"equivalent modes of expression of the same or similar semantic concepts which may be related to each other through simple structural transformations" <page 49>.

Schwarcz considers attempts at modelling the extra-linguistic environment as either unrealistic (that is, non-representative), or impractical. He suggests that instead, the conceptual structure associated with a given input sentence be presented with it. One problem with this form of extra-linguistic feedback is that the designer of such a system runs the risk of designing his conceptual representations in such a way that they are inadequate for complete language representation. The model will, however, appear to acquire the language because it is learning ways of associating two patterns, the input sentence and the associated conceptual structure. Though this defect will, of course, show up eventually when the model is unable to perform well in ordinary discourse, it seems preferable to avoid the pitfall by supplying the model with a non-linguistic environment which is a good model of the world independent of linguistic interpretations.

The most important capability of the model, according to Schwarcz, is that of inductive generalization. This capacity enables the modification of the syntacto-semantic component by:

> "(1) the formation of classes in order to combine several rules into a single rule, (2) the addition of new items to these classes, (3) the inference of inclusion relations between, and ultimate

> merging of, classes specified in different rules, and (4) the induction of transformational equivalences among different rules." <page 50>

Schwarcz gives the rules by which each of these actions would be accomplished, in general terms. Two other forms of induction are

> "(1) the segmentation and classification of utterances into morphemes and (2) the application of both paradigmatic and syntagmatic rules of conceptual inference to produce changes in the conceptual network" <page 50>.

Conceptual representations within the model are graph structures built from units "of the form R(a,b), where R is a relation or operator, and a and b can be anything at all" <page 45>. Some of these elements are innate, or "paradigmatic"; others are learned, or "syntagmatic".

As Schwarcz points out, this "triadic" representation is common in many existing natural language processing systems (see Chapter 4). He suggests that these structures could be extended to syntactic structures, and that the process of comprehension can be modelled by a sequence of transformations of a surface graph structure of lexical items into a conceptual graph structure. This structure can then interact with the semantic base of conceptual graph structures in order to elicit the appropriate response. Production of utterances is seen as the inverse of comprehension, starting with a conceptual substructure and ending up with a a graph of lexical items that is converted

to a "string of phonemic or alphanumeric symbols, which is output by the processor" <page 47>.

Schwarcz, like the authors of the previously described models, has left out many details necessary for an implementation of the model. However, what he <u>has</u> done is to attempt to <u>identify all the components</u> necessary for a performance model of the typical speaker, and to outline <u>specific classes of methods</u> which, on the basis of some previous success, might prove to be practical tools for implementing the model. Secondly, he has attempted to define, in a way which is useful for a computer implementation, the stages through which acquisition goes, with respect to the <u>structures which are formed and the environmental data used to form them</u>. Whether he is right or wrong is not as important as the fact that he has offered a paradigm which has some possibility of being used.

3

Modelling versus Pragmatism

3.1 Methodology

The term "artificial intelligence" implies a dichotomy. Intelligence is to be replicated, hence we are modelling a characteristic of human beings; the goal is, however, an artifice, so it will in some way differ from the original and give in to practicability. Similarly a language acquisition system is an attempt on the one hand to explain the rise of the child to linguistic competence[1], on the other to increase the efficacy of natural language processing systems.

The value of a computer system to linguistic theory development is dependent on the attempted degree of isomorphism between the implementation and the child. For example, consider the following extreme on this scale of isomorphism. The model of the child is a "black box"

---

[1] Henceforth I shall not ascribe Chomsky's technical meaning to this word.

program; its inputs are identical to the child's, and so are its outputs. Then the only correspondence induced by the model is child<-->program, which is not terribly interesting.

Faithful representation in an acquisition model must be tempered by the limitations of hardware, expense, and the state of software implementation of related psychological mechanisms, such as perceptual processes and semantic representations. Furthermore, the other goal of acquisition programs is to give such natural language processing tools as general question-answering (Q-A) systems and perhaps mechanical translation (MT) programs the dimension of adaptibility. Such systems are constrained by sub-goals which are different from those of the human. For instance, Q-A systems may be required to be oriented to numerical computation or factual retrieval from huge data bases. They may need a simple perceptual component, their environment simply being a data base. MT programs may have no environment at all, and may learn through feedback in the target language, a technique which has little in common with the way children learn <Dale 1972>. In the following sections we will examine possible characteristics of a natural language acquisition system and their potential variation between faithful representation and practicability.

3.1.1 Linguistic input

The linguistic experience of the child would most accurately be replicated by direct acoustic input of speech. However the problem of processing acoustic signals is far from solved <Reddy 1971; Tappert, Dixon 1973>. D.R. Reddy *et alia* <1973> have written a system which attempts to use interaction between the phonological, syntactic, and semantic components to understand statements of chess moves. The acoustic component does a partial harmonic analysis of the continuous speech input and hypothesizes phoneme boundaries. Word recognition is done by starting with a hypothesis from the syntactic component, and passing these hypotheses back and forth among the three components until all but one are disconfirmed. The disambiguation procedures are *ad hoc* and highly dependent on characteristics of the game of chess, and the vocabulary is very small, so it is an open question whether it is possible to generalize the methods to larger and more comprehensive systems.

An alternative to acoustic input is a character encoding of input, using phonetic symbols. Subtle gradations of breath, pitch, stress, and idiosyncratic articulation would be lost, and hence cues which the child has available would not be replicated. For instance, children may utilize the change in intonation between adult-child and adult-adult conversation as a cue to ignore ensuing discourse <Kelley 1967>. The other drawback to this

method is that it would be clumsy and would require a highly skilled trainer. If an acquisition system is to be useful, it must be usable by a relatively untrained person. Similar objections can be made to the use of phonemic transcriptions, for which the loss of cues would be even greater.

The most widely available convenient input form is conventional orthography. Every general-purpose computer system has means for accepting input encoded in an orthography which is usable by a large community of people, and conventional orthography eliminates many dialectal variations.

The difference between printed language and the child's linguistic experience is considerable. More cues would be lost than in a phonetic or phonemic encoding, but in addition the segmentation rules to be learned would be different. Written word boundaries are clearly marked in most languages, and there would be no problems with perceptiblity or articulation. Homophones would be replaced by homographs, and morphemes would contain different numbers and distributions of allomorphs. For instance, the English regular plural morpheme might be represented by _s_ and _es_, as opposed to /s/, /z/, and /əz/. For these reasons, generalizations which are difficult for the child may be easier for the machine model, and _vice versa_. Orthography is thus the pragmatic extreme on the modelling<-->pragmatism

scale, given that the input form must be a natural representation of language (that is, ruling out linguistic pre-processing by the user, such as bracketing constituents, _et cetera_.)

3.1.2 Linguistic representations

The choice of mechanisms and strategies to be incorporated in the acquisition system will logically determine formal structures which will be built up for the comprehension and production of utterances <Schwarcz 1967>. For instance, if simple associative bonds are the only tools available, production will presumably be a matter of chaining elements of associative networks together to form utterances. If, on the other hand, the system can hypothesize phrase-structure rules, it will develop patterns for choosing and invoking these rules.

If it were easy to observe or infer which mechanisms the child actually uses, we could incorporate models of them in the system and observe what structures are produced. However, because adult linguistic behaviour has been far easier to observe than the child's acquisition methods, representations of adult speech and adult performance are far better developed than models of acquisition. Hence it is easier to allow a chosen representation of adult speech to constrain the mechanisms of the acquisition model.

Because of this "cart before the horse" phenomenon, the choice of linguistic theory which will underlie the parsing and generating components will be very important. If "quick and dirty" parsing is desired, and semantic disambiguation is unimportant, then a context-free grammar may be sufficient; efficient, generalized context-free parsers are well-known, but incorporation of categories based on semantic features into a context-free grammar induces huge inefficiencies. A transformational-generative grammar might be used if the system learned only by generating sentences, or might be used for the generative component while another representation is used by the parsing component. A case grammar might be chosen for generation because it is more compatible with the semantic component. Clearly there are many dimensions to the modelling<-->pragmatism scale as exhibited by the choice of linguistic representation.

3.1.3 Non-linguistic environment

As with linguistic input, the most realistic extra-linguistic input method would be artificial sensory inputs, such as TV cameras, microphones, and thermocouples, and the disadvantages are similar. While visual pattern-recognition techniques are fairly reliable in a simple geometric environment, they require expensive hardware and a large share of computer resources and are not

yet suitable for more complex environments in which texture, colour, and amorphous objects are important <Agin and Binford 1973; Slobel 1973>. If the system has a variety of perceptual inputs the problem of interaction has to be dealt with. Interaction between hand and eye has been modelled with some success <Feldman and Sproull 1971; Feldman et al. 1971>, but nothing has been done about audio-visual interaction, for instance.

Limiting perception to the visual reduces processing problems, but limits the experiential context within which linguistic interchange and acquisition can take place. Isomorphism between the model's and the child's experiences is then restricted to a portion of the child's experience. Concepts associated with weight, temperature, <u>et cetera</u> would have to be learned, if at all, through linguistic experience only. In effect, they would become more abstract than the same concepts for a child. Hence differences would be expected between the progress of linguistic expression of these concepts by the child and by the model.

The environment may be dynamic; that is, elements may be allowed to change position, disappear, reappear, <u>et cetera</u> Tracking of movement, however, is at present a difficult, expensive activity in terms of computer resources, so that it would be more practical to modify the representation of the scene after each change in

configuration. This kind of strategy assumes that changes are discrete and separated by sufficient real time to allow such modifications.

There are also alternatives for interaction among human, model, and environment. The child is exposed very early to a number of cues which differentially reinforce both linguistic and non-linguistic behaviour. If the model is able to act upon its environment, and if in addition the human can selectively reward appropriate non-linguistic responses to commands, important information is then available to modify parsing procedures. Although this type of information is available to the child, it is unknown whether he uses it. Appropriate non-linguistic responses are also useful for evaluation of the success of acquisition.

Reward for appropriate linguistic responses might enhance the rate of acquisition, but there is no evidence for this for children. Ervin-Tripp <1971> and Dale <1972> have pointed out that parental attitudes to children's utterances depend on the truth value of the utterances, not on their syntactic well-formedness. There are some exceptions: profanity, and constructions from a dialect which the parents disapprove of. The value, to learning, of rewarding linguistic output is unknown.

Closely related to perception is motor ability, including ability to move about the environment and

manipulate objects in it. These abilities allow the system to learn concepts directly instead of by abstract definition in terms of other concepts. In order for non-stative verbs to be learned, there must be either ability to act on the environment or else arbitrary motion of objects in the environment. Purposive verbs could be learned as expressing elements of the model's physical problem-solving abilities.

The ability to move about the environment would enable the system to learn the intrinsic and extrinsic meanings of words like "right", "left", "front", "back", et cetera. Expression of concepts involving weight, mobility, animacy, speed, and time might be learned in a way more akin to the child's if the concepts were actually an integral part of the system, not defined in language.

The cheap and easy way to provide an environment to which linguistic input will refer is to simulate it using the computer. An imitation world which appears on a CRT can quite easily be programmed <Winograd 1971> to the level of complexity desired. The simpler the world model, the weaker the isomorphism with the child's non-linguistic experience. In the extreme, the non-linguistic input could be unrelated to that of the normal child. For instance, it could be an encoding of a circuit-diagram, data structures, or arbitrary 2-dimensional patterns. These forms could in fact be those which are important to specific applications.

3.1.3 Non-linguistic environment

3.1.4 Cognition

A language acquisition system without cognitive ability would bear almost no relation to the child. Wheatley <1970> has used the term "engagement" to describe the link between utterances and the state of the speaker and his environment at the time[1]. It is highly likely that a system without cognitive ability would at best learn a taxonomic description of the input language, since it would have no independent criteria for assigning interpretations to utterances. Ervin-Tripp <1971> cites some evidence that lack of non-linguistic input linked directly to linguistic input inhibits language learning. Hearing children of a pair of deaf parents watched a great deal of television, but by three years of age could still not comprehend or produce speech. This could be due to the lack of correspondence between utterances from the television and phenomena in their personal environment.

The level of current machine models of cognition is such that some similarity with the child might be expected. Winograd's <1971> BLOCKS world is a simple model of a 3-dimensional world which contains objects with properties like shape, colour, et cetera. The scene could easily be

---

[1] "...to discover the engagement rules it is presumably necessary to inspect more than just well-formed sentences of the language; it would seem to be a good guess that one must also understand something of the forms of social behaviour. ...there seems to be no obvious way to detect engagement rules mechanically." <Wheatley 1970, page 36>

modified to embody some other concepts which the child might have.

To make the cognitive ability even more representative, it could be made dynamic, that is, learning of concepts could be taking place at the same time as learning of language. There is strong evidence that cognitive development is one of the major determiners of the course of acquisition in the child <Slobin 1971b; Bowerman 1973; Brown 1973>. Unfortunately, the models of cognitive learning are far less developed than pre-programmed cognitive models, so in order to concentrate on the language acquisition problem, a relatively static cognitive component is probably more practical.

3.2 Criteria for success

To be related to the child, the acquisition model must have the ability to answer questions and produce spontaneous speech (spontaneous in that it will not be in response to linguistic input). In a system with the ability to affect the environment, it should be able to carry out actions in response to linguistic input.

The objective evaluation of performance in these domains is difficult. There are two basic ways to monitor performance. One is to examine the content of the data structures built by the model over time, and the other is to

conduct linguistic tests analogous to those used to test humans. The first runs the risk of being too subjective, since it is up to the designer or someone familiar with the workings of the program to decide on the status of the vocabulary and parsing and generating rules. The second method may be satisfactory, and would be a good topic for further investigation both by AI workers and linguists.

3.3 Existing systems

In the following sections most of the existing computer systems for learning natural language are reviewed. Analysis of the systems will refer in part to the characteristics described in sections 3.1 and 3.2.

3.3.1 Dependency grammar

McConlogue and Simmons <1965> have written a pattern-learning parser (PLP) which attempts to learn how to correctly construct dependency analyses of English sentences. A dependency analysis of a sentence is a tree whose nodes are the words of the sentence. The offspring of a node are said to be "dependent" on that node, and the node itself "governs" its offspring. A dependency grammar is essentially a set of rules which assigns words to word-classes and which specifies how combinations of

word-class symbols occurring in a sentence induce governor-dependent relationships among words of those classes. Figure 3.1 shows a dependency analysis of an English sentence.

The learning sequence is a fairly simple one. A sentence is presented to the system. The system assigns it a dependency analysis based on its previous experience with the words in the sentence. If it is then told that it is in error and the correct analysis (produced by a human) is presented to it, it revises its dictionary and grammar rules according to the correct analysis.

The PLP was subjected to a number of experiments. It was given 300 sentences in basic English in the learning mode. (That is, it was presented with the correct analysis for comparison after each attempt.) Its success rate followed a typical learning curve, with 91 of the last 100 analyzed correctly. After these it was given a further 100 sentences from the same source text as the first 300 in both the learning and non-learning modes. With no further learning it analyzed 77 correctly; with learning, 88. Furthermore, it was able, on a second pass, to analyze all the first 300 sentences correctly.

Overall, the PLP is not designed to resemble a child acquiring a first language. First in importance is the fact that it has no environment. That is, there is no language-independent component against which the PLP can

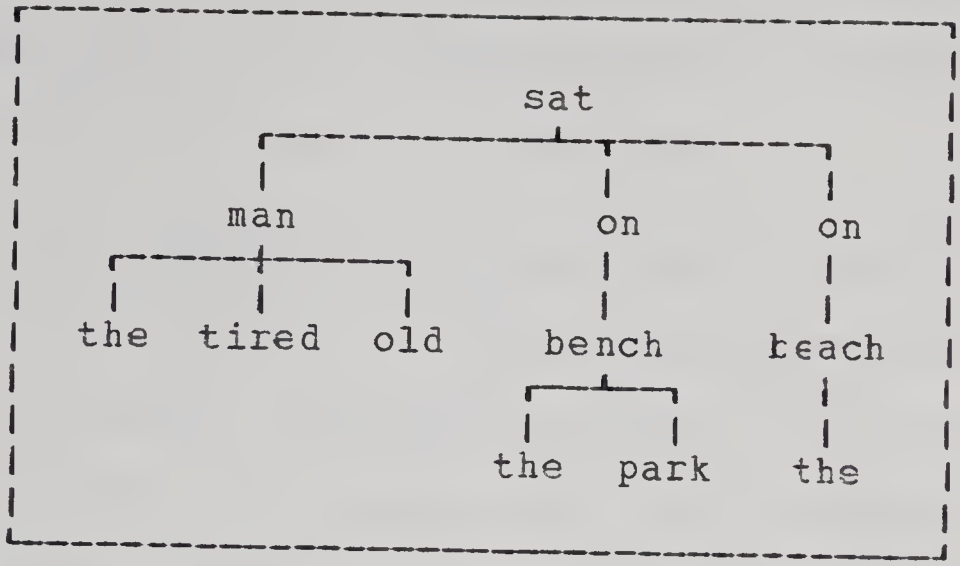


Figure 3.1 A dependency analysis of "The tired old man sat on the park bench on the beach." <McConlogue, Simmons 1965, page 689>

match incoming sentences. This lacuna may be consequence of the lack of any explicit semantic interpretation of dependency analyses in general. From McConlogue and Simmons' description, a dependency grammar seems to be a purely taxonomic description of a language. Terms such as noun phrase, verb phrase, etc., are used in the dictionary phrase-structure rules, but there is no explicit independent motivation for such categories and no specification of their semantic significance.

A second fault of the dependency grammar used by the PLP is that it does not do any morphemic analysis. This is actually consistent with its lack of semantic interpretation, since there is no need to relate "see", "saw", "seeing", for example, to a common meaning if meaning is unimportant to the system. However, this means that many of the generalizations which have to be learned by the child

are irrelevant to the PLP. Furthermore, the multiplication of dictionary entries for related forms of the same stem is hardly tolerable in a practical system.

Because no semantic interpretations are available, the PLP must rely on feedback of correct dependency analyses. This kind of feedback represents another major difference between the system and a child, for the child never sees correct analyses of utterances to which he is exposed, except for rare instances in later schooling, well past the major stages of primary acquisition.

The PLP is not meant to model the child, and hence does not have any method for generating sentences or other responses to linguistic input.

Linguistic input is, as expected, conventional orthography. However the need to input complex grammatical analyses in a difficult metalanguage makes the system impractical for general use.

Because of the feedback of correct dependency analyses, the PLP can be judged for success purely on the basis of its analyses, compared with the correct hand analysis. Hence the evaluative experiments outlined by McConlogue are simple and objective.

The main value of the PLP to a more comprehensive language acquisition system may be in the designing of first-level strategies for the preliminary analysis of input

utterances. A partial dependency analysis may allow a first guess at a grouping of constituents without any semantic analysis. This preliminary "chunking" could be passed on to the more general semanto-syntactic analyzer, which would try this preliminary grouping first.

3.3.2 Hypothesis-testing

Kelley <1967> has designed and implemented a system which attempts to model the early stages of the acquisition of language by the child. The fundamental hypotheses inherent in the model are that:

(1) syntax is learned by testing hypotheses about the language which are influenced both by innate characteristics and by the particular language to be acquired;

(2) confirming and infirming hypotheses depends crucially on semantic interpretation; and

(3) utterances must be partially understood in order to be used for hypothesis-testing.

Unlike the PLP described in section 3.3.1, Kelley's system "serves as a hypothetical model of [syntactic acquisition]." <page v>

Input to the program is, as usual, orthographic. The sentences are generated randomly from the target grammar by a routine in the system. The linguistic representation

constraining the types of hypotheses tested is a context-free phrase-structure grammar. However, the parser has two novel features. First, it uses weights which have been assigned to the phrase-structure rules to calculate weights for competing analyses of the same sentence. The second, more important feature is that it can construct analyses which use only part of the input sentence. Thus if a sentence cannot be given a full structural description because parts of it use some unknown rules, the parser will skip those parts and analyze the rest if it can.

The course of acquisition in the model is broken into three stages, the beginning of each of which is marked by the arbitrary creation of new "initial hypotheses" which are intended to be extralinguistic in origin (that is, innate or learned independent of language).

The only initial hypothesis in Stage 1 is that a sentence consists of one word which belongs to the lexical category "thing" and has the function "concrete reference of the sentence". During Stage 1 the program randomly selects a word from each sentence input and assigns it to the "thing" category and gives it the function "concrete reference". As it does this it continually increments confirmation of the Stage 1 initial hypothesis, since this is the _only_ hypothesis, and since analysis with this hypothesis is _necessarily_ correct.

In Stage 2 a new category called "action" and a

function "modifier of the sentence" are added. This means that eight new hypotheses are possible, and these are all added to the hypothesis set. The strongly entrenched hypothesis from Stage 1 will still be available for use in an analysis, but since analyses are highly weighted by the proportion of the utterance for which they account, it will be less significant. Where it will be significant is in biassing analyses towards assigning, to words which have already been placed in the "thing" category, the function of concrete reference.

At this point an important question might be asked. Since random words encountered in Stage 1 have been assigned to the "thing" category, then some verbs must be in that category. How will they get moved into the "action" category? That is, how will the program know that it has made an incorrect analysis in analyzing an incoming verb as "thing" and "concrete reference"?

The answer is that there is a comparator component which has access to a very important piece of information: the correct parse of the incoming utterance. Since the program has no environment or even a semantic base, Kelley has substituted a kind of semantic match of each putative analysis with the correct analysis. In Stage 1 this match is not done, but in Stage 2 the constraint on an analysis is that "the main verb of the correct analysis may not function as the [putative] 'concrete reference' of a sentence nor may

it be [putatively] considered a member of the 'thing' category" <page 141>. The importance of the comparator component should not be underestimated. Without it there would be no way of choosing between putative analyses or of confirming and infirming hypotheses. The ccmparison constraints represent an attempt to formalize the notion that the child can build up a consistent system for understanding language only if he can assign to an utterance an interpretation in which some aspects of the semantic representation within the speaker are reproduced.

It is my contenticn that for the child in the early stages, the link between him and the adult that enables him to assign crudely similar semantic interpretations is the envircnmental ccntext in which the utterance takes place (see Chapter 4). Since Kelley's system has no model of such an environment, it must revert to the artifice described above. The problem with this heuristic is that it uses terms which are grammatical in nature to connect the correct and putative semantic interpretations. This criticism really applies to Stage 3, in which the functional relation "subject of the sentence" is introduced. Kelley states that "The semantic ccntent of this relation is taken to be quite similar to what the relation of the same name means in an adult grammar" <page 122>. However, it is hardly an established fact that "subject of the sentence" is a semantic relation at all. Case grammar thecries indicate that the grammatical, or surface subject can be derived from

almost any deep structure case. Case grammar is a universally accepted theory of neither semantics nor syntax. However it shares with most of the recent experimental semantic processing systems a desire to formulate some primitives which can be utilized to describe events and relationships in the real world. So when, for Stage 3, Kelley adds to his comparator the condition that "the head of the phrase that functions as subject in the correct analysis must be the same word that functions as subject in the putative analysis" <page 141>, he is really making the comparison heavily dependent on _syntactic_, not _semantic_ information. It is highly unlikely that the child has any access to this kind of information.

Six possible patterns of functional relationships are used to formulate initial hypotheses of Stage 3. In addition, although there are no new lexical categories, the two-word patterns of Stage 2 are considered as possible categories which might be constituents of the three-word patterns implied by the three functional relations of Stage 3. Thus any of these categories can _a priori_ serve any of the three functions. By the end of Stage 3 (that is, after 180 sentences have been processed), the subject-predicate hypothesis is well-confirmed. That is, the rules

S -> THING *3

*3 -> ACTION THING

are well-confirmed, and *3 may be thought of as analogous to a verb phrase.

The source of the lexical categories and functional relations at the beginning of each stage is not made clear. Presumably in the child they occur as a result of increased cognitive abilities. Hence the separation of stages in this way is a more practical way of simulating dynamic cognitive abilities without having to program those abilities or an environment within which cognitive abilities could be learned.

Kelley has combined the two methods of evaluating success of the model. He advances the entrenchment of the internal subject-predicate pattern after 180 sentences as evidence that the system is progressing, like the child, towards the target grammar. Furthermore, he has experimented with data which contains ungrammatical combinations of constituents, and has found that the same subject-predicate pattern was learned, and that the categorization of lexical entries was converging on those acquired with the first set of data.

Kelley's system is a valuable step towards a more comprehensive model of human acquisition and a viable practical acquisition system. The principle (rather than a particular implementation) of a comparator component which evaluates the semantic acceptability of a putative analysis is likely to be an important part of any acquisition system. The weighting of hypotheses according to their successful use in parsing seems to be a fruitful method for confirming

grammatical rules. Also useful may be the parsing method which is able to make a partial analysis of a sentence, for it seems clear that the child must do this <Wheatley 1970> in order to understand utterances the complete rules for which he does not yet control.

3.3.3 Associative networks

Jordan <1972> has created a system called METQA (Mechanical Translator and Question-Answerer) which she claims contains no pre-programmed linguistic or logical ability, and yet will learn to understand any orthographic language, translate from one language to another, and answer questions from its acquired knowledge. She has attempted to do this by designing the system around a primitive associative net.

The net is built from nodes of the following types:

terminal node: a node through which the net may be entered or left. Each terminal node is connected by a *description* link to its *description* node.

description node: a node containing a string of characters (usually a word) which has occurred as a segment in some previously experienced string

idea node: a node that has *transform* links to all

terminal nodes in an equivalence class (e.g. "dog", "chien", "hund", etc.), and *combination* links to all *fact* nodes using that idea node

fact node: a node which has *combination* links with *idea* nodes and *fact* nodes

p-rule node: a node which describes, using *context classes*, how to permute string segments.

In addition, any node N may have *class* links to nodes which represent the classes to which N belongs.

The labels on the directed edges of the network are described as follows:

transform link: a link which is part of a path joining words which are "equivalent" (i.e. which are allowable substitutes for each other). This link is usually undirected.

combination link: a link connecting idea and fact nodes to fact nodes

description link: a link connecting a fact node to an idea node or a terminal node to its description node

class link: a link connecting an idea node N to a node representing a class to which N belongs

membership/subset link: a link connecting a node N to a node which N contains as a member or subset

p-rule link: a link connect a node N to a p-rule node which applies to contexts involving N

<u>equivalence link</u>: a link connecting nodes which are mutually substitutable

In addition, any link except a description link can be <u>restricted</u> by a context requirement list which specifies class requirements for surrounding words before this link can be traversed. Each link is also weighted.

Clearly, in terms of the characteristics outlined in section 3.1, we have very little to evaluate. There is no identifiable internal linguistic representation, no non-linguistic environment, and no cognitive or motor ability. Then how does METQA get feedback to learn? The answer is that, unlike the child, it must get explicit linguistic feedback from the trainer. In fact, Jordan maintains that "There is no direct attempt made here to simulate the human mind, either in the form of semantic memory network used or in the manner in which METQA learns" <page 22>. What Jordan is trying to do, then, is to see just how far an associative network pattern-recognition approach to language learning can be pushed, without attempting to model the human. The questions then are, "How does it work and how well does it work?"

In order to learn how to translate from one language to another, METQA is given a sequence of pairs of strings. The first member of each pair is a substring S1 of a sentence in language L1, the second S2, its correct translation into language L2. METQA transforms S1 into a putative

translation T2. It then attempts to transform S2, the feedback string, to S1, and adjusts weights on certain links depending on how close the S1->T2 paths were to the S2->S1 paths.

METQA uses a fairly sophisticated pattern-matching method to segment the input string and produce a set of putative matches of segments with terminal nodes in the memory net. The "only requirement for its input is that it be a string of discrete symbols" <page 21>. This implies that a standard orthographic, phonetic, or phonemic input form could be used. Also, in principle, morphemic analysis is an integral part of the processing, unlike any of the other acquisition systems reviewed here.

Once all the matching terminal nodes are identified, METQA chooses the set of non-overlapping terminal nodes which leaves the least amount of the input string unidentified. This is called a <u>cover</u> of the input string. It then proceeds with a breadth-first traversal of the network. The checks made before choosing a link to traverse from each node are:

(1) Is this a transformation of the feedback string? If so, is the next link a part of the original transformation path?

(2) Is the next node connected to a p-rule node which the current input string must satisfy?

(3) Which of the possible next links is most highly weighted?

In addition, if a traversed node has a p-rule attached, then the current context is checked, and the p-rule applied, if possible.

The basic learning mechanism is as follows. If, while processing the feedback string S2, a path intersects the S1->T2 path on a node N <u>other</u> than a terminal node, then it adjusts the weights on the links leading from N. A link from N which was used during S1->T2 but not from S2->S1 is <u>downweighted</u>, along with its context rules, since it was a "bad" choice. The link G by which the S2->S1 path arrived at N is <u>upweighted</u>, along with its context rules. Also, if G has context class rules, then the segments of S1 are assigned membership in those context classes.

Of course, not all segments of S2 will find intersections with S1->T2 paths as they are transformed through the net. These segments must somehow be matched with unmatched segments in S1. METQA uses a variety of heuristics, including splitting and permuting segments, to assign putative matches and context restrictions.

The way in which METQA learns p-rules and context classes makes the role of the trainer very important:

> "To the extent that the trainer understands METQA's procedures, observes METQA's behavior, and infers what METQA has already learned, then he can present permutations in a sequence which will produce more general and intuitively pleasing

> p-rules and classes, and <u>fewer</u> of them, than if permutations are learned from a haphazard input sequence" <page 107>.

Jordan seems to think that the child receives similar training: "...just as the child is molded by his environment and benifits from wise teaching, the program METQA is strongly influenced by its human teacher" <page 143>. This is a surprising statement in view of the fact that children are <u>not</u> taught how to understand and speak their first language.

METQA can also operate in a fact-assimilating mode and a question-answering mode, but since they have no effect on the learning process, I will not discuss them here.

Jordan presents the output of a training run with METQA as evidence of its success, but the results are far from impressive. She concentrates heavily on translation as a feature of the system, and in fact this is the only feature implemented. It is unlikely that learning to translate from a string in one language to a string in another is similar to learning either language. The published results show only two examples of sentence translation (from French to English) and one of them is incorrect.

It is questionable whether Jordan has succeeded in her intent to eliminate pre-programmed linguistic and logical abilities. She has attempted to define a network with a very simple set of primitives (that is, a few types of nodes, a small set of relations, and simple integer

weighting), but has replaced linguistic and logical structure of the usual type with an extremely complex control program. Despite the fact that they superficially do not resemble ordinary logical and linguistic structures, the algorithms by which the data structure is manipulated do some of the same jobs as the logical and linguistic components of more conventional systems. For instance, the control program explicitly creates subset-superset relations among nodes, and constructs permutation rules for reordering string segments. It also has special predefined input symbols to tell it whether the input is a question, and two predefined classes of input words: question pronouns and negation words.

The lack of a non-linguistic environment obviously denies METQA the status of a model. However it also seems that lack of non-linguistic feedback makes METQA's acquisition process inefficient and perhaps inadequate. Independent semantic input would allow the system to organize class relationships far more easily and less arbitrarily than by strictly linguistic context.

Jordan uses feedback in a second language to teach translation ability, but there is no evidence that there is any way for the system to learn a single language. Feedback in the same language would teach the ability to paraphrase, perhaps, but this bears little resemblance to language understanding. The training schedule is so important that

it is possible that it is really the trainer who is learning METQA, not METQA learning language!

These criticisms raise an interesting conjecture. Suppose that a non-linguistic environment were added to METQA in such a way that the region of the environment relevant to an input string could be focussed on. The encoding of the focal region as a subgraph of the memory network would be considered as S2, the feedback "string". The conjecture is that Jordan's network traversal approach could be applied to transform the input string towards the representation of the focal region, producing a subgraph which would be a "parse" of the input string. Transforming the correct focal region subgraph back towards the linguistic terminal nodes could result in the same kind of learning techniques METQA now uses, but it would be making use of language-independent information to create classes, nodes, links, etc., as well as linguistic information.

Jordan's experiment is a useful one, and, as pointed out above, the network concept might well be used as a basis for an acquisition system, given some changes and the addition of some components. In particular, the pattern-recognition method for segmenting input strings might provide a general approach to morphemic analysis.

3.3.4 A language-acquiring robot

The most ambitious attempt at a comprehensive language acquisition system so far is Harris' <1972> robot system. He has attempted to include in his system a number of the components and relationships hypothesized for the child and his environment.

The robot's linguistic environment is divided into three phases, only the last of which is at all like the environment of the child. Although Harris claims to be "trying to mimic the conditions under which a child learns a natural language" <page 85>, Phase 1 consists of a sequence of (word-list)<->(concept-list) pairs, a kind of input the child never has. A "concept" is a name of a physical or mental capability of the robot, like its motor or its pathfinding algorithm.

There are situations for the child which might seem similar to Phase 1. For instance, suppose an adult somehow directs the child's attention towards a cat and says "Pussy, pussy." This seems like a pairing of word with concept. But the child may include in his focus of attention many objects, many attributes of those objects, and many categorizations of those objects and attributes. He might associate "pussy" with the concepts <u>rug</u>, <u>red</u> (colour of the rug), <u>grey</u> (colour of the cat), <u>move</u> (what the cat is doing), <u>small</u> (relative to other animate objects), <u>animal</u>

(the category of animate non-humans), _et cetera_. Because the adult cannot inject the concept _cat_ into the child's brain the child must await further experience to disambiguate the many possible pairings of word and concept. Harris seems to place undue emphasis on this direct word<->concept type of learning, considering that the only direct analogy he draws with the child's experience is an example like the "pussy" one above <page 90>.

Another concession to practicality (or ease of programming) is the pre-processing of idioms and special word-combinations like "left of", "right of", "_il_ y a", _et cetera_. In the typed input, the component words are connected by underscores and the system treats the combinations as single words. This is both unrepresentative of natural language and of dubious value to more general language acquisition. The system is in principle unable to learn generalizations of "left of" to "top of", "back of", "side of", _et cetera_, and to connect "il y a" with its inflections for aspect and tense as in "il n'y a pas", "y a-t-il", "il y avait", _et cetera_.

The method of establishing correlations between words and concepts seems successful in this case, and it might be useful in the early stages of any acquisition model.

Phase 2 represents the largest discrepancy between the child and the robot. Input at this stage again consists of ordered pairs. The first component of each pair is an

English sentence (actually pseudo-English, for example, "A chair is right_of the table.") The second component is a list of the "parts of speech" of the corresponding words in the first component. The "parts of speech" are categories of the concepts built into the robot. For instance, concept 1 is the motor and belongs to the part of speech "V". Hence the robot can immediately connect each word it sees in phase 2 with a part of speech. This information is something the child *never* gets in the whole course of acquisition.

Harris' system is a good example of what happens when the designer feeds in what what he thinks are the semantic representations of utterances. The risk is that he will define semantics in such a way that it is hardly different from syntax, and hence he is really feeding in high-level syntactic information. This gives the grammar-inferrer an easy task, but makes the process less realistic and useful.

Since the purpose of Phase 2 is to infer a grammar, this is an opportune point at which to describe the linguistic representation used by Harris. Based on the internal division of processes, we can separate the representation into syntactic and semantic components. The syntactic representation of the target language is a "transformational context-free grammar" <page 83>. This description is misleading, however. The grammar inferred bears little resemblance to Chomskian transformational-generative grammars or any of their variants.

The grammar contains a set of context-free (CF) phrase-structure rules of the form

A->a

A->BC...Z

A->B

A->aB

where $a \neq \emptyset$. These rules generate the <u>surface structure</u> of a sentence. An additional constraint implicitly imposed by Harris is that there be three productions of the form

S->.

S->?

S->!

where S is the start symbol. The starting production is chosen according to the terminating punctuation of the input sentence.

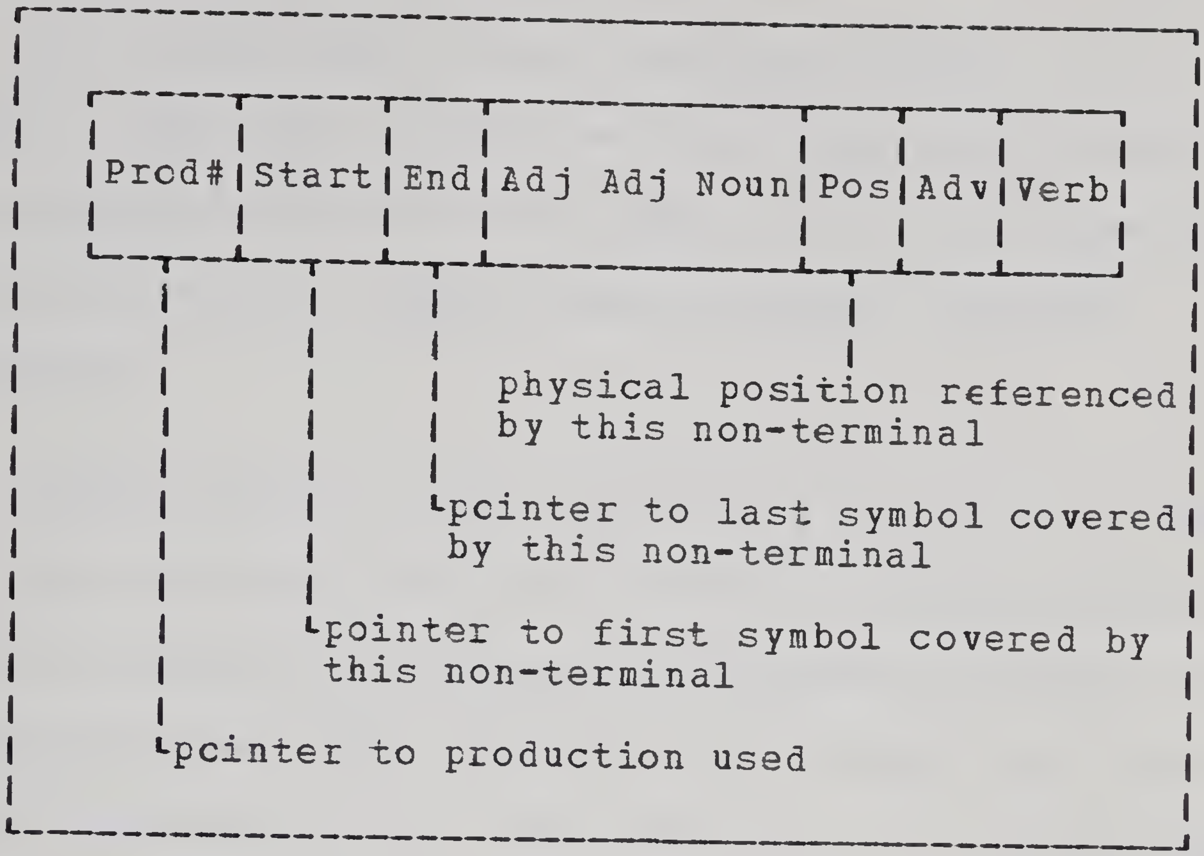
In addition, the grammar may contain reordering transformations relating <u>surface structures to deep structures</u>. These transformations reorder the constituents immediately dominated by ".", "?", or "!". Since these are the only transformations, and their inverses and the deep structure phrase-structure rules are easy to compute, the grammar can just as easily be used for generation as for parsing.

The semantic interpretation of a sentence is generated by application of rules which are associated with the form of the CF productions. As productions are applied in the

top-down parse of an input sentence, a list of empty semantic descriptors is built, one descriptor for each instance of a production. The form of a descriptor is shown in figure 3.2. When a non-terminal is completely expanded as a series of terminals, the semantic routines associated with the productions are backed up its subtree.

For productions of the form A->a, the concept corresponding to "a" (in the lexicon) is inserted in the corresponding descriptor. For productions of the form A->BC, the descriptors for B and C are combined, so that the descriptor of A is the same as that of C except for cells for which C is empty and B is not, in which case the cell from B is used. If the resulting descriptor for A is self-contradictory, the representation of the environment (that is, the semantic base) is consulted to resolve the contradiction. Although Harris doesn't mention it, the ordering of B and C makes the semantic component language-dependent, since it is designed specifically to take care of, among other things, English relative clauses, and hence would not be applicable to languages in which modifying complex constituents preceded the head word.

Productions of the form A->B simply back the semantic descriptor for B up to A. The most significant semantic routine is the one for productions of the form A->aB. For this, the routine calls the "concept routine" for the concept associated (in the lexicon built in Phase 1) with



<u>Figure 3.2</u> The form of a "semantic descriptor"

the terminal "a". These specialized routines are pre-programmed, and they are dependent upon the "meaning" of the concept. They alter the descriptor associated with B in a specific way and copy it into the descriptor corresponding to A.

The result of backing up all these descriptors is a single descriptor which is associated with the whole sentence. Clearly, from the form of the descriptor, it is unlikely that this kind of representation is adequate for a much wider class of sentences than those used by Harris. For instance, no sentence with compound constituents could be represented by such a descriptor. Sentences with conditional clauses or adverbial clauses would be

unrepresentable, for example

If the big chair moves, go to the pianc.

Without gcing intc a detailed characterization of the limitations of Harris' descriptor, it is clear that it is insufficient for a large class of natural language sentences.

We now return to the workings of Phase 2. A grammar of the fcrm described above is inferred by constructing, on the basis of the set cf sentences experienced in Phase 2, a trivial grammar which generates all and only the "part of speech" patterns given. For instance, for the patterns "ab." and "bc?" (these are artificial examples), it would generate the grammar G:

```
.->12
?->23
1->a
2->b
3->c
```

It now applies, in order, if it can, three cperators to the grammar G. These operators are such that their resulting grammars describe languages which contain L(G) as a subset. They are applied as follows.

<u>grouping</u>: Group the two non-terminals which appear in sequence most often in the starting productions, and replace them with a new non-terminal to generate them.

<u>folding</u>: If there are two productions of the form

A->B and C->B

or cf the form

A->BM and C->BN,

then replace all occurrences of C in the grammar with A.

recursion: Replace any non-terminal that appears repeated (like C in A->BCCD) with a new recursively generable non-terminal.

A characteristic of the grammar inferred by these operators is that they may generate sentences which have forms which are different from those of the original set of sentences. This, of course, is desirable in a model, but might, unless constrained, get out of hand. The reasons why Harris' implementation does not are clear. When the grammar is used for parsing, it doesn't matter if the parser is too general, since it is always presented with correct sentences. When the grammar is used for generating replies, it is constrained by the pre-ordained form of the semantic descriptor which is the origin of the reply.

These phenomena may be representative of the child if we accept that comprehension and production are fairly separate processes. However, it is difficult to tell whether, if Harris' system were given a more comprehensive semantic interpretation component, overgeneralized grammars would produce grossly aberrant utterances.

Harris has developed his own heuristic parsing method because he claims that one cannot know ahead of time whether the CF grammar is LR(k), SLR(1), or any other type covered

by existing generalized parsers, and that there is no parser for general CF grammars. However, the Earley <1966> parser is a generalized CF parser, and it would be interesting to compare it with the "bandwidth heuristic search" algorithm of Harris.

The non-linguistic environment of the robot is a map of a rectangular grid, in which there are several objects, each coded by a single concept number and a pair of coordinates. In addition, the robot's parts are part of the environment; for example, the motor (1), the robot itself (2), and its contact device (6).

The robot's cognitive abilities are embodied in the semantic routines associated with concepts. These routines seem only to be able to handle concepts of motion and position, both relative and absolute. In this respect the use of adjectives like "big", "little", "small", "green", _et cetera_, by the robot is misleading, for the robot has no ability to assign independently-motivated meaning to these words. That is, they are simply labels which are attached to the concepts representing the physical objects. There is no colour other than green, and the robot has no eye, so green is just a label. There is no evidence of any categorization of the words "big", "little", and "small" which would relate them to dimensions or even to each other.

Harris' criterion for success is a subjective assessment of the responses of the robot to a sequence of

sentences entered in Phase 3. We can assume that sentences listed are all different from those used in Phase 2, and all, with the exception of one, are sentences which are generable by the grammars induced in Phase 2. It is difficult to tell without seeing all the Phase 2 sentences whether there was significant generalization of structures in Phase 3.

As far as the quality of responses goes, the robot performed quite well, at least for those sentences presented in the thesis. Restatements of declaratives were consistent, answers to questions were correct and natural, and non-linguistic responses to commands were appropriate. As with all such subjective programmer-generated tests, however, one is always left with a nagging suspicion that there are sentences which would foul up the robot, not by being unparsable, but by misleading the semantic routines.

A more substantive criticism of the robot's performance is that it fails to partially understand sentences which its grammar cannot handle. It is clear that children, and probably adults as well, can assign partial semantic interpretations to sentences whose complete structure they cannot analyze, and this capability seems possible to implement, judging from Kelley's <1967> experience (see section 3.3.2).

Although Harris claims to model the conditions surrounding a child's acquisition process, I have pointed

out a number of glaring dissimilarities, including the direct (concept-list)<-->(word-list) pairing of Phase 1 and the "part of speech" patterns given for sentences in Phase 2. Though it must be granted that simplification is necessary in an experimental system, the simplification of cognitive abilities and environment in the system seems to have trivialized some of the semantic capabilities of the robot, notably the interpretation of adjectives. Furthermore, the semantic routines are specific to each concept, and do not seem to be generalizable to a more comprehensive set of concepts.

Except for the _ad hoc_ and highly constrained nature of the semantic descriptors, the comprehension system seems fairly successful. One must keep in mind, however, the large amount of syntactic information the adaptive routines were given with which to infer a grammar, and the extensive pre-programming of the semantic routines.

The lack of any morphemic analysis is bound to induce inefficiencies in any large-scale implementation of this system, for reasons explained in section 3.3.1.

As a practical system, Harris'system has the drawback that the teacher must have a knowledge of the structure of the target language, in terms of the robot's "parts of speech". In fact, it is difficult, from Harris'evidence, to deduce how much the teacher's skill influences the success of the system.

On the positive side, as far as the development of a natural language acquisition system goes, Harris' system has many good overall features. It embodies components which are necessary for any successful system: an environment[1], cognitive capabilities[1], a parser-constructor, a generative component, a semantic base[1], and a semantic interpretation component[1]. In addition, the semantic interpreter guides the parse, and the robot and the human can both act on the environment. For these reasons it is a very useful first step, and should be examined by anyone who tries to construct a natural language acquisition system.

---

[1] This component is not contained in the other natural language acquisition systems reviewed here.

4

CLAP: a Paradigm

4.1 Introduction

The purpose of this chapter is to describe the characteristics of CLAP, a comprehensive language acquisition program. The description will be concerned with possible learning schedules, internal organization of the system into components, and possible ways of implementing each component. Division of CLAP into components does not necessarily mean that the ideal system would be thus organized, but it enables us to separate areas where solutions do not exist from those where existing techniques and representations can be applied.

One phenomenon which has retarded the progress of mechanical understanding of natural language is the continual desire to re-invent the wheel. Each researcher seems to think that *his* graph structure is going to be better than all the other graph structures, or that his parser will handle more English than the previous ones. Too

seldom have workers in AI taken an existing representation or technique and attempted to push it as far as possible towards the solution of a specific problem. For example, who has picked up SHRDLU <Winograd 1971> and tried to make it understand conditionals, complicated time references, or multiple nominal groups? If the task of modelling the language acquisition process can be broken down into subtasks, we can hope that techniques drawn from existing AI systems can be used in experiments with these individual components.

4.2 The learning schedule

It is very important, before describing the internal organization of CLAP, to specify the alternative classes of input sequences which could be used. The criteria for classifying such sequences are:

1. Is the linguistic input of a kind matched to the current stage of acquisition or does it always consist of natural adult utterances?
2. Is the non-linguistic input artificially constrained and constructed so that minimum ambiguity and hence maximum speed of learning is achieved, or is it more like the ambiguous, multisensory type that children experience?
3. Does the Human give explicit feedback to the

system as to the nature of its errors and successes or is the system dependent solely on non-linguistic information and its own conceptual system for comparison and evaluation?

As with all existing acquisition systems, linguistic input to CLAP will be orthographic, for reasons outlined in section 3.1.1. The linguistic input should not be designed to suit the current stage of acquisition, since human trainers would vary widely in their skill at matching input to the sophistication of CLAP, and children do not seem to _require_ such register shifts, though adults often use them.

Within the limits of the non-linguistic environment chosen for CLAP, non-linguistic input will be natural. If the environment is a CRT display, then regions will be pointed to, windowed, moved, or otherwise transformed, by the Human and by CLAP. If the environment is a real-life scene and CLAP has perceptual apparatus like a TV camera, then the Human will be able to move and focus the camera, change lighting, manipulate objects, etc. It would be unrealistic to input events in the form of internal encodings of special concepts which could not be otherwise communicated by the Human; Schwarcz <1967> suggested this and Harris <1972> implemented such a scheme (see sections 2.3.5 and 3.3.4).

The form of feedback will vary with the progression of

acquisition (see the following sections). In the early period, there is simply development of associations between words (or, more generally, substrings) and concepts. Later, when CLAP starts to produce utterances, feedback will come mainly from non-linguistic input and special approval-disapproval input, seldom from linguistic input. Still later, when much fuller understanding of sentences is achieved, human linguistic responses will make up a large part of the feedback on which additional learning will be built. At no stage, however, will there be explicit feedback, by the Human, of correct parses. An exception would be corrections stated in natural language; such corrections occur (though relatively rarely) in parental speech to children.

Given the above constraints, each input by the Human will be one of the following types:

- u - an utterance, in orthographic form, including punctuation;
- a - an action on the Environment, including moving objects, changing lighting, or focussing on a region of the Environment;
- r - a coded input indicating approval, disapproval, or lack of understanding;
- s - a stimulus to output spontaneously;

or any unordered combination of the form: (u,a), (r,u), (r,a), or (r,u,a). Combinations involving s do not occur, since each of the other types of input is assumed to be a

stimulus to output.

CLAP may respond in kind, except that r can only be an expression of non-understanding. (There's an ethical red herring here: why can't CLAP disapprove of the Human's actions? But that's another story!)

The introduction of responses invites the question of how the Human responds to CLAP's responses; an even more fundamental one is what CLAP does with the Human's responses. Are they interpreted as rewards, or in some other way linked to CLAP's performance, linguistic or otherwise?

The Human's response to utterances should be as natural as possible. If the utterance seems to be declarative, then the Human's next input may or may not be related to it. If it seems to be interrogative, he should answer; if imperative, he should follow the request. CLAP can interpret the Human's responses as rewards (that is, as satisfying the goals of the Action-Taker[1]) depending on the content of those responses. Human responses to actions may be coded stimuli representing approval or disapproval (that is, satisfying or not satisfying the goals of the Action-Taker). They may also, however, be linguistic responses. The sequence of events is that represented by figure 4.1.

---

[1]I will elaborate on the Action-Taker's role in responses in the following sections.

Human input	Possible CLAP output
u	u,a,r,s,(u,a)
a	u,s
r	u,a,s,(u,a)
s	u,a,s,(u,a)

CLAP output	Possible Human input
u	u,a,r,s,(u,a),(r,u),(r,a),(r,u,a)
a	u,a,r,s,(u,a),(r,u),(r,a),(r,u,a)
r	u,a,s,(u,a)
s	u,a,r,s,(u,a),(r,u),(r,a),(r,u,a)

Figure 4.1 Conversational sequences with CLAP

4.3 Strategies

The learning strategies which CLAP brings to bear on its experience will vary with time, with each strategy acting on the results of the previous one. A similar temporal dependency probably exists in the child, but it is further complicated by the effect of cognitive development. Though the effect of this development is extremely important, I will not consider it here.

Schwarcz <1967> has outlined five identifiable stages in language acquisition (see section 2.3.5); they include some of Brown's <1973> stages but do not induce as fine a subdivision as his. The interesting thing about these and other hypothesized acquisition stages is that they are based

almost exclusively on observation of child utterances, and are hence expressed in terms of categorizations of those utterances and an identification of the sequence of emergence of those categories. This direct dependence on the classification of utterances is a reflection of the obsessive preoccupation of linguists with production over comprehension both before and after the transformational-generative reformation. I will propose an alternative method for defining the stages of acquisition.

CLAP's acquisition process is divided into stages only implicitly on the basis of its learning strategies - not the learning of production, but that of comprehension. As we shall see below, the governing process is the building of the Parser. Hence production phenomena occur as a result of structures which have been built in the Parser. For example, when linear order regularities appear in CLAP's output, it is because linear ordering has already been comprehended by the Parser to an extent which allows the Responder-Modifier to construct corresponding generative mechanisms.

Figure 4.2 represents the organization of CLAP into control components and data components. Clearly this separation into component types is arbitrary; it is also useful. The feedback loop of the Human, Parser, Evaluator and Parser-Modifier, Action-Taker, and Responder is shown; however, the Parser-Modifier and Responder-Modifier may also

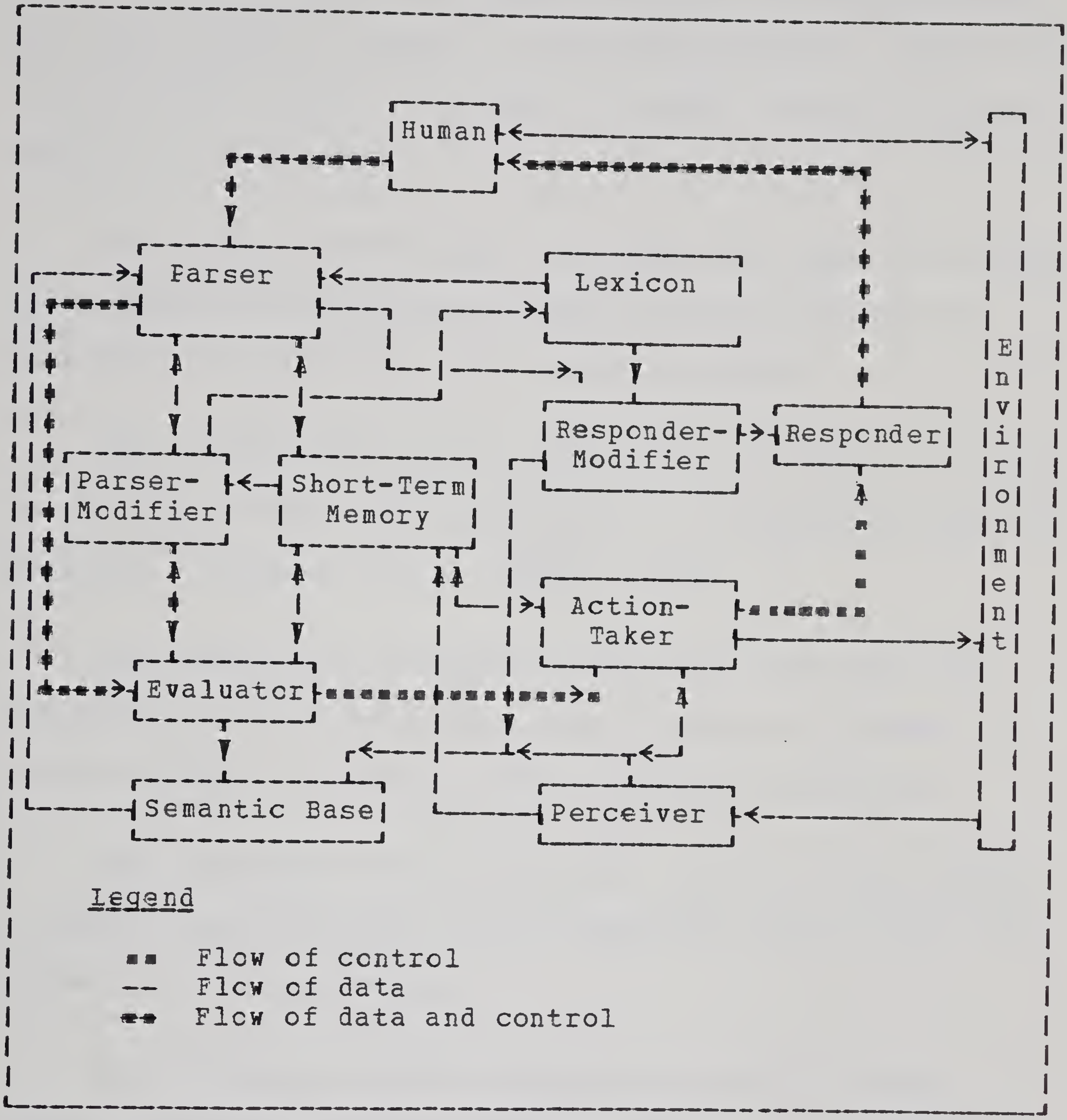


Figure 4.2: The components of CLAP

operate asynchronously.

The Parser contains a control structure which will use the Lexicon and a set of Focal Structures obtained from the Short-Term Memory to segment the incoming Utterance, and to produce a semantic structure, the Parse. The Evaluator

looks at the Parse and attempts to assess its plausibility. This evaluation is passed to the Parser-Modifier, which uses it in deciding which structures to modify and how to change weights in the Parser.

The Parser-Modifier uses the Parse and Focal Structures (both stored in Short-Term Memory) to change the Parser, extending its structures or modifying weights.

The Action-Taker can use the Parse to perform an action in the Environment, add information to the Semantic Base, or initiate a response to the Human's input.

The Responder is a control structure which uses the Intention, constructed by the Action-Taker, to attempt to build a string of segments which is CLAP's utterance.

The Responder-Modifier uses the Parser and the Lexicon to add to the structures in the Responder and to modify the weights on its components.

The Perceiver receives input representing changes and/or events in the Environment. Changes are translated, if necessary, into changes in the Semantic Base, and events are stored in the Short-Term Memory.

The following sections will expand the above sketches of the functions of CLAP's components, and section 4.4 will offer suggestions as to their implementation.

4.3.1 Strategy 1: segmentation and meaning association

The functions of Strategy 1 are assigned to two different developmental stages by Schwarcz <1967> (see section 2.3.5). However, for CLAP, there are two motives for combining the functions. The first has to do with the way the child learns segmentation. Without entering the debate over whether the child learns phonemes, syllables, or any other formal grouping of phones, we can say that he learns to recognize certain repeated chunks of the acoustic input stream. In other words, he learns some boundary markers in incoming speech. If these boundaries are to be of any use to him in later assigning meaning to utterances, then sooner or later (it must be before his first utterance) he must choose to segment at least some of the time at meaningful morpheme boundaries. Now we can regard this process of choosing meaningful morphophonemic boundaries as analogous to the search for meaningful morphographemic boundaries, and this must of needs be accompanied by some process of association of orthographic segments with concepts. Hence Schwarcz' first two stages are combined. What we are saying is that if there is indeed a preceding stage in the child in which he learns a purely phonetic characterization of the corpus, then that is at best analogous to the division of input text into characters, and hence is irrelevant to CLAP.

The second consideration in combining the two stages is

pragmatism. It seems to be far easier to learn morphemic analysis with the help of the Environment and the Semantic Base than without.

The object of Strategy 1 is to discover a way of breaking down an incoming sentence, including blanks and other punctuation, into segments which can be associated with concepts. This breakdown should be such that, in some future sentence, the associated concepts can be put together to form a coherent interpretation when the sentence is segmented in the same way.

For example, suppose the sentence

The blue pyramid is big.

is segmented

|T|he_|bl|u|e|_py|ram|id_|is_|big|.|

and the segment "big" is associated (by the procedure outlined below) with the concept #RELSIZE. Then if a future sentence

Pick up the big pyramid.

is segmented

|Pi|ck_|up_|t|he_|big|_py|ram|id|.|

then at least part of its interpretation of the sentence will be correct. On the other hand, if the segment "bl" is associated with the concept #BLUE, and a future sentence

Pick up the black box.

is segmented

|Pi|ck_|up|_t|he_|bl|ack_|bo|x|.|

then CIAP has a problem.

Olivier <1968> has written a program which uses a stochastic mechanism to assign probabilities to segments in a dictionary, revising the probabilities on the basis of "stretches" of 480 characters. However, Olivier's problem was somewhat different from the one posed here. His program was given compressed text, that is, text whose blanks and other punctuation have been deleted. The goal was to learn to segment the text into its original words. CLAP, on the other hand, is trying to divide each sentence into segments each of which is a clue to the meaning of the sentence. Hence, it is trying to discover morphemes in the text.

Figure 4.3 shows the elements of the Parser and of CLAP which are involved in parsing before Strategy 2 has come into effect. CLAP starts with a Lexicon consisting of the set of characters which can appear as input. The first sentence will be segmented into single-character segments:

|T|h|e|_|b|l|u|e|_|p|y|r|a|m|i|d|_|i|s|_|b|i|g|.|

The Parser-Modifier then forms a new segment from each pair of adjacent segments:

(Th he e_ _b bl lu ue e_ _p py yr ra am mi id d_
_i is s_ _b bi ig g.).

A set of concepts is associated with each segment appearing in the sentence, in the following way.

Concurrent with the entering of the sentence, the Human

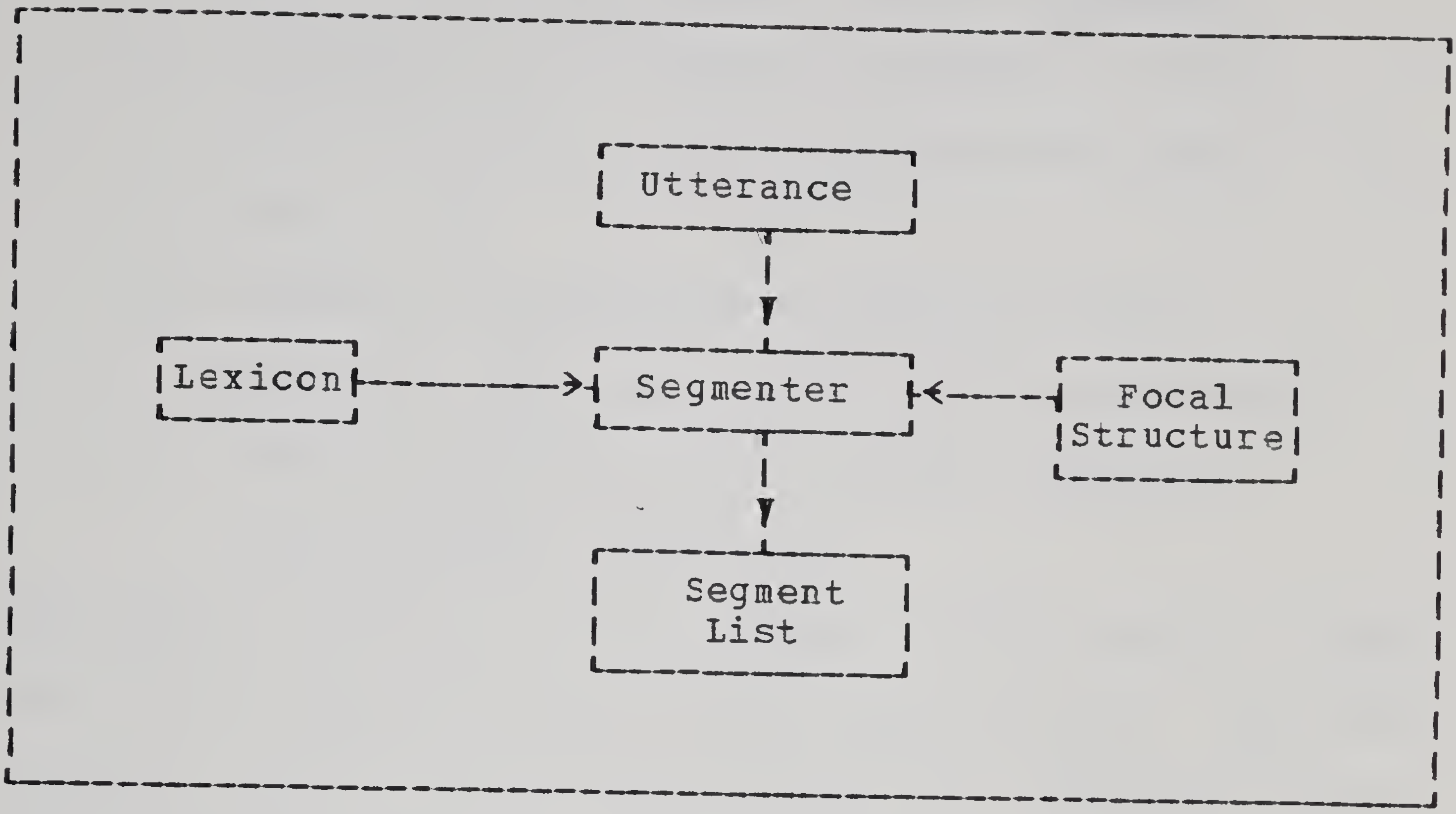


Figure 4.3: Elements involved in parsing under Strategy 1

enters information by means of the Environment[1]. This information consists of any changes the Human wishes to make to the Environment, and/or a directing of CIAP's attention toward some portion of the Environment.

CIAP's Perceiver component analyzes the Focal Region in the following way. It creates a list of pointers to all the concepts which are associated with the Focal Region. These will include:

1. elements of the Semantic Base which represent referents in the Environment,
2. elements of the Semantic Base which are abstract concepts, such as classes and

[1]We will assume, for the moment, that the Environment is a scene displayed on a CRT.

relations among objects, events, *et cetera*,

3. elements of the cognitive component, such as those involved in problem-solving and plan-making, and
4. elements representing the ability of the system to affect the world, such as components which manipulate objects and the component which produces utterances.

The concepts available to be pointed to will depend a great deal on the sophistication of the Semantic Base. There are numerous criteria which could be used for choosing concepts for the Focal Structure. For an example of one heuristic, see section 5.2.3.

Having thus formed a sub-environment, or Focal Structure, the Parser-Modifier will attach to each of the one- or two-character segments produced by the parse a pointer to each of the concepts in the Focus, and will attach an initial weight to each pointer.

For the second and subsequent sentences, the initial parse into segments goes differently. As the parse (we will call segmentation "parsing" since it is the first stage of any full parse of an incoming sentence) proceeds from left to right, segments are chosen according to a static evaluation measure which involves the relative frequency of the putative segment and the degree to which it has been clearly associated with a concept in the current Focus.

The resulting segmentation, or Parse, can now be interpreted by creating a list of the concepts with which the segments have been highly associated in the past. The Evaluator compares the interpretation with the concepts directly associated with the current Focus. The results of this comparison are passed to the Parser-Modifier, and those concepts in the interpretation which do not appear in the Focus have their association weights lowered. All concepts in the Focus are then associated with the sentence segments as in the first case, and their weights are augmented.

After segmentation new segments are created, as before, by the Parser-Modifier, by combining adjacent segments and adding them to the Lexicon.

As more sentences are processed, the Lexicon will come to consist of a list of segments, each of which has associated with it a list of weighted pointers to concepts. Clearly there must be an ageing process for useless segments. The measure of utility of a segment involves the time since last use and frequency of usage, so this information must be attached to each entry.

There is, of course, no _a priori_ guarantee that the above procedure will result in perfect learning of morphographemic analysis. However, the success of Olivier's <1968> program in segmenting compressed text into words leads me to believe that it has an excellent chance of

providing enough segmenting skill that reasonable interpretations can be assigned to new utterances. Olivier's program suffered under handicaps which CLAP does not have: compressed text having no blanks or other punctuation; arbitrary "stretches" of text instead of relatively homogeneous sentences; and no evaluation of "meaningfulness" of segments.

The Action-Taker's job in Strategy 1 is, of course, dependent on the input. If the input is an utterance, then the Parser will, if it can, produce an interpretation consisting of one or more concepts which are clearly associated with a segment or segments of the incoming utterance. These concepts may be any of the following:

1. an object in the Environment;
2. an attribute of an object in the Environment, such as its colour or shape;
3. a relationship between two concepts in the Semantic Base, such as #SUPPORT or #COLOR (an attribute relates an object to the value of the attribute);
4. a procedure for manipulating the Environment;
5. a procedure for problem-solving;
6. a procedure for output-generation.

Since, in Strategy 1, the Parser can only present the Action-Taker with an unstructured list, the best the Action-Taker can do is examine the concepts on the list and see if any of them represents a completely specified procedure. If

one of them is found, the Action-Taker may invoke that procedure.

The Action-Taker will do more than this, however. It must contain some global goals, in order to simulate some of the human incentive to acquire language. Among these goals should be:

1. the desire to accumulate knowledge;
2. the desire to communicate knowledge to the Human; and
3. the desire for approval from the Human.

Thus, in examining the output from the Parser, the Action-Taker is looking for

1. something it can add to the Semantic Base;
2. something indicating a desire for CLAP's knowledge;
3. approval.

During Strategy 1, the Action-Taker is unable to add to the Semantic Base because the concepts produced by the Parser are not explicitly related. One of the concepts in the Parse may be an output procedure, but since such an output-procedure will probably require specific information, and since the concepts supplied to the Action-Taker by the Evaluator are not explicitly related, the Action-Taker will not be able to invoke the output procedure. Hence there will be no replies to utterances during Strategy 1.

Actions are a different matter. When the Human inputs

an action, CLAP has a semantic structure to work with. If, for instance, the Human moves an object in the Environment, CLAP focuses on the origin and destination of the move, and creates a conceptual structure representing this Focal Region. The Action-Taker can take the structure, select, for instance, the top level, and pass it to the Responder, which will attempt to construct an utterance based on the current state of the Grammar. In Strategy 1, this corresponds to selecting those concepts with which the Responder-Modifier has associated lexical items. This list (possibly null, unary, or longer) is output. Initially there will be many null or unary utterances, because of the primitive state of the Lexicon. The output at this stage will model the first single-word utterances of the child.

If the Human input is disapproval, the situation is more complicated. Short-Term Memory contains a record of the last utterance by CLAP, so a trace of its production is available. The only thing CLAP can do is to diminish the weight on the links between the words in its utterances and the concepts with which they are associated. This diminution will be related to the degree of confidence CLAP has in its weights. That is, well-established links should be decreased less than poorly-established ones. Approval by the Human should result in a strengthening of the weights on the word-concept link.

Input of a simple stimulus, s, is simply a request for

a response by CLAP. In this case, CLAP will respond in the same way as it does to an action by the Human, except that its Focal Region will be determined by some heuristic. One alternative is to use the last Focus; that is, to allow CLAP's area of attention to remain unchanged.

To sum up, then, Strategy 1's job is to build the Lexicon and the simple weighted links between the Lexicon and the Semantic Base. It needs the Environment in order to pass some kind of information between the Human and the Semantic Base, and it needs text from which to derive segments for the Lexicon. It has not finished its job when Strategy 2 begins.

4.3.2 Strategy 2: linear ordering

As soon as the Lexicon has developed to the stage where CLAP can understand more than one word (segment) in an incoming utterance it can begin to attach import to word order. However, this is not the only phenomenon of Strategy 2 - just the obvious one. Strategy 2 is really concerned with building structures from a few morphemes. It happens that when these structure-building routines are used to *generate* utterances, word-order regularities appear.

In order to construct a Parser, there must be recognizable building blocks. We have already seen one such block: the weighted link between segments and concepts.

Strategy 2 produces the second major building block, and produces it as soon as CLAP starts to build the Structure-Builder (see figure 4.4). The Structure-Builder is, or rather is built as, an Augmented Recursive Transition Network (ARTRAN) of the type that Woods <1969> has described. However, I will not restrict myself to the specific representation used by Woods, nor will I assume that the network can only build transformational-generative deep-structure trees. In fact, such networks can be used to build any finite structure, including those used for semantic representations, as exemplified by Winograd <1971> and Simmons and Slocum <1972>.

We have seen the operation of the Segmenter in Strategy 1. In Strategy 2 the Segmenter does the same job. The Concentrator examines the Lexicon, the Segment List, and Focal Structure to determine a Target Structure, a semantic structure with the following properties:

1. it appears in the Focal Structure,
2. it contains all the concepts which are clearly associated with segments in the Segment List, and
3. it is the smallest and most highly-weighted structure which satisfies conditions 1 and 2.

Property 2 above raises a question: when is a concept "clearly associated" with a segment? Unfortunately, it is difficult to answer the question without some data on the kinds of associations built up between segments and concepts

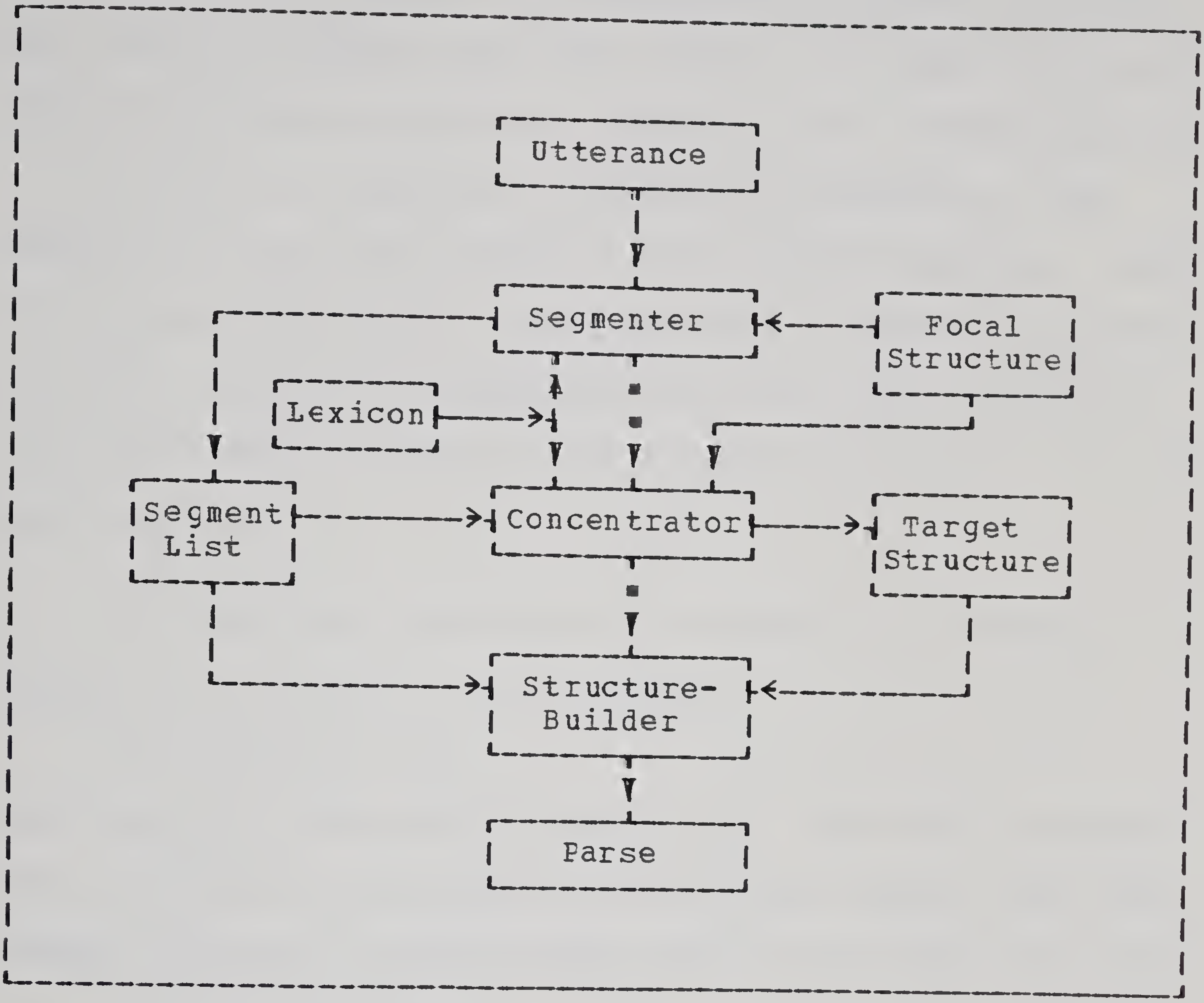


Figure 4.4: Elements involved in parsing uncer Strategy 2

by Strategy 1. Fcr this reason, it is probably advisable to implement pilct programs to investigate each Strategy in the crder in which they are described here. Initially, of course, the Structure-Builder is empty, since the Parser-Modifier has not had a chance to build anything. Thus there is initially no Parse tc pass to the Parser-Modifier. Instead, the Target-Structure is passed to the Parser-Modifier.

The Parser-Mcdifier now builds the first part of what

is to become the Parser. It examines the Target Structure and builds an ARTRAN which will accept the Segment List and output the Target Structure. There will not necessarily be a one-to-one correspondence between the segments in the Segment List and the concepts in the Target Structure. Each arc in the ARTRAN will be weighted with a number calculated using the likelihood associated with the Target Structure and the relative weights of the concepts associated with the input segments.

To clarify the process, let us consider an example. Suppose CIAP receives the utterance

P1 is on top of B3.

and the focus indicated in figure 4.5. Schwa The Perceiver will produce a Focal Structure which will perhaps look like (using Winograd's representation plus a few added concepts)

```
((#IS :P1 #PYRAMID)
 (#IS :B3 #BLOCK)
 (#IS :B2 #BLOCK)
 (#IS :TABLE #TABLE)
 (#COLOR :P1 #GREEN)
          •
          •
          •
 (#SHAPE :B3 #RECTANGULAR)
          •
          •
          •
 (#FRONT :B4 :P1)
          •
          •
          •
 (#IS #RED #COLOR)...)
```

Let us assume that the Segmenter produces the Segment List

(P1 _is _on_ top_ of_ B3 .)

and that the segments "P1" and "B3" have been clearly

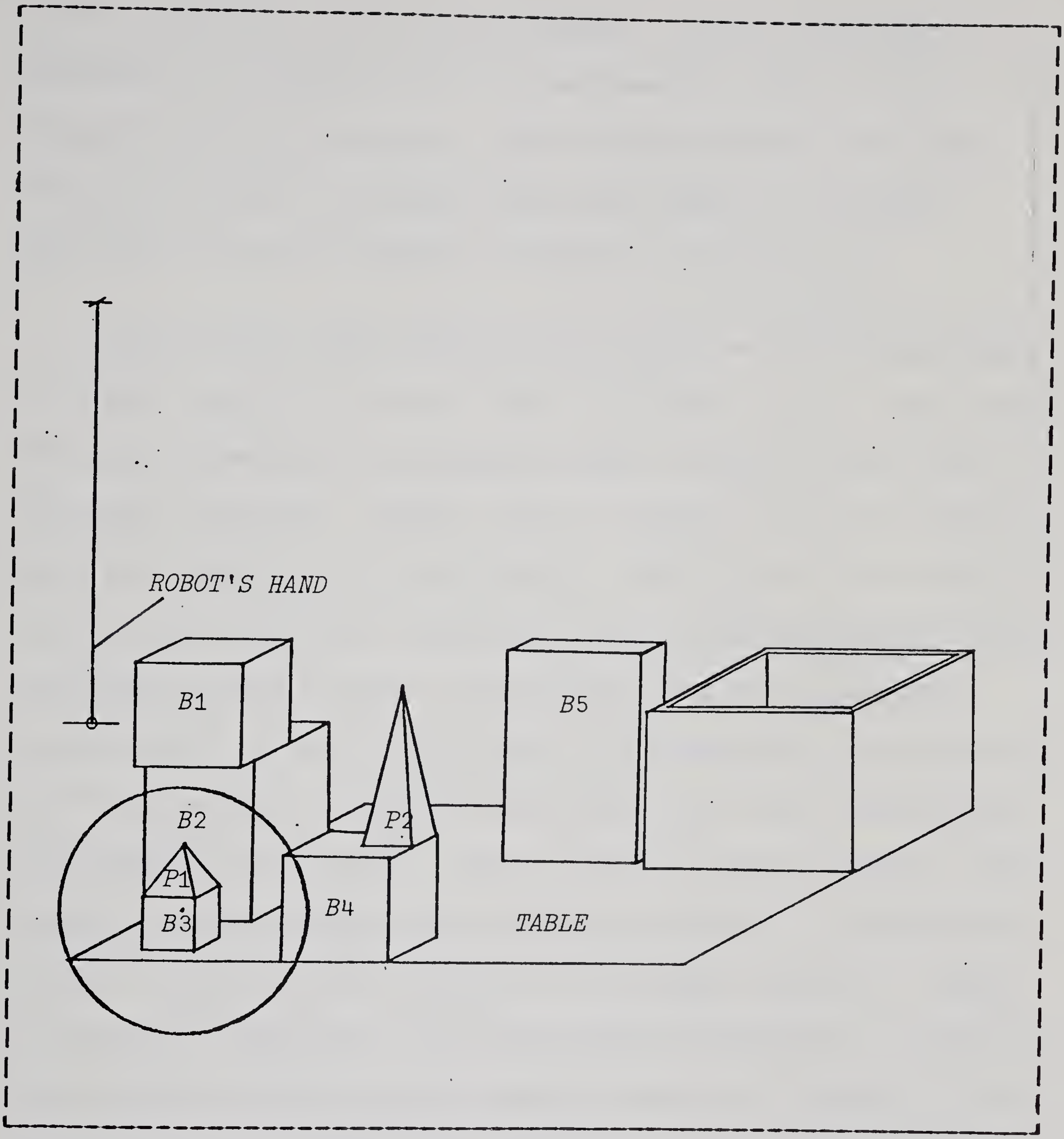


Figure 4.5: The Focus input with "P1 is on top of B3."

associated with the concepts :P1 and :B3. Then the Concentrator will have to choose as Target Structure one of

```
(#IS :P1 #PYRAMID)
(#IS :B3 #BLOCK)
(#COLOR :B3 #RED)
(#SUPPORT :B3 :P1)
(#COLOR :P1 #GREEN)
```

Since (#SUPPORT :B3 :P1) contains 2 concepts linked to words

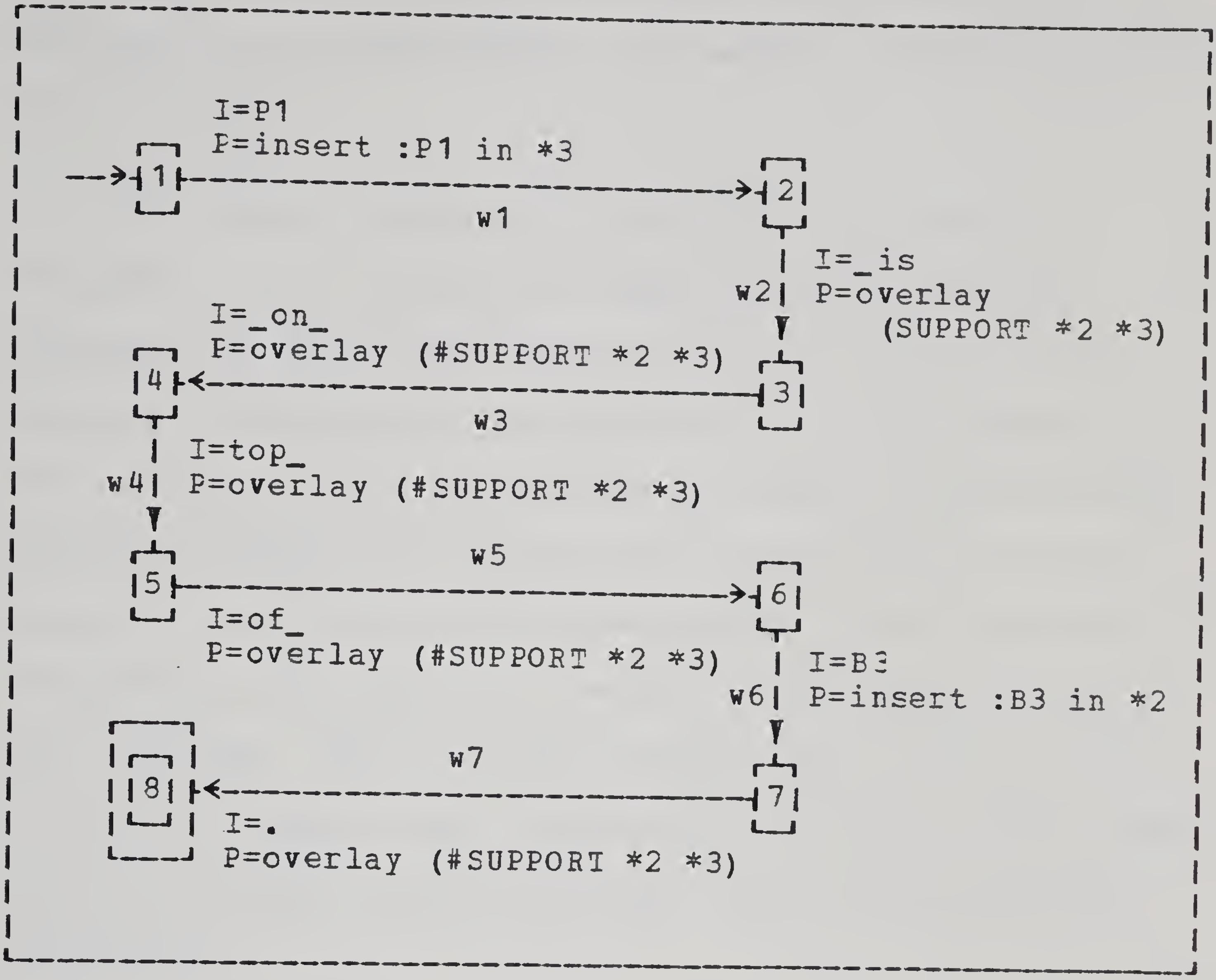
in the utterance, it will be chosen. Since the Target Structure is already part of the Semantic Base, it is accepted by the Evaluator. The Parser-Modifier will now attempt to build an ARTRAN which can create the Target Structure, with the result pictured in figure 4.6.

It is clear that the only part of the Parse generated by known[1] segments are the arcs (1,2) and (6,7). The other parts are generated by segments most of which do not have clear-cut meanings. There will be several types of unknown segments occurring in the ARTRAN. One type will be nouns, verbs, and adjectives (generally open-class morphemes) whose reference is not clearly discernible in the Environment. Another will be what I will call the "modulating morphemes". The "modulating morphemes" are those that Brown has called "grammatical morphemes". They include aspect markers, tense markers, prepositions, case markers, articles, auxiliaries, copula, person, person markers, and number markers. (Many of these are expressed by closed-class morphemes.) A list of some English modulating morphemes appears in figure 2.2, and section 2.2 discusses their appearance in children's speech.

Brown's term is inappropriate because it obscures the important semantic role of these morphemes. There may be some morphemes learned at this stage which are purely syntactic, like gender inflection for inanimate nouns, but

---

[1]A "known" segment is one which is clearly associateed with a concept.



<u>Figure 4.6:</u> A first ARTRAN (For an explanation of the notation, see Appendix 4.)

in general this is not so. It is difficult to characterize these morphemes precisely. At best, we can say that they have no directly observable referent in the Environment, they modulate the meaning of an utterance, and their meaning lies in relationships among the referents of the segments learned in Strategy 1.

A prerequisite for the learning of meanings of the modulating morphemes is that they have been entered in the Lexicon in Strategy 1. They will have survived in the

Lexicon not on the strength of their association with referents in the Environment, but because of their frequency of occurrence.

The general strategy for learning the modulating morphemes is the following. When the Parser-Modifier examines the Target-Structure it will find parts of the structure which are not generated by the known segments in the utterance. For instance, in the sample utterance which generated figure 4.5, though "_is" and "_on_" are in the Lexicon, they have no clear-cut meaning. The occurrence of #SUPPORT in the ARTRAN is a result of the co-occurrence of "P1" and "B3". The arc which accepts "_on_" may in future be used to analyze other occurrences of "_on_", and in cases where it has the locative meaning, the analysis will be appropriate.

Clearly these new nodes and arcs are not always correct, but the augmentation and diminution of weights over time and the removal of low-weighted arcs will eventually result in a Structure-Builder which can generate correct Parses in many cases.

To indicate the possible kinds of nodes to be introduced in Strategy 2, we will examine Brown's <1973> list of English morphemes (see figure 2.2), and hypothesize the kinds of nodes which will be built.

Present Progressive. Two nodes are involved: one to

accept the auxiliary, one to accept the suffix. The first will not build any structure. The second will build (or fill in) concepts representing the duration and relative temporal position of the event being constructed.

"<u>in,on</u>". Brown means the locatives here. Since "in" and "on" occur with many other meanings, there will be many other (possibly wrong) nodes built for them. However, two which will survive will be nodes building or filling in the structures (#CONTAIN <object1> <object2>) and (#SUPPORT <object1> <object2>).

<u>Plural</u>. The plural markers have many meanings, like "in" and "on". In many cases, number is marked redundantly, since the modifying adjectives provide sufficient evidence from which to infer number:

two red blocks
some pyramids
other colours.

The structures onto which the plural marker will be mapped will depend on the general method of semantic representation used in CLAP. Further, they depend on the structures created by the Perceiver. One of the initial assumptions about CLAP was that the cognitive and perceptual mechanisms would be static. Thus it will be necessary to pre-program the Perceiver with faculties which are mature enough that CLAP will have access to a sufficient range of structures to build a rich Parser. Hence the Perceiver will attempt to create a Focal Structure containing quantification markers

explicitly rather than implicitly.

Past tense. As mentioned above, the Focal Structure is actually a set of structures representing past states and events back to the limit of CLAP's Short-Term Memory. If the utterance is related to a past event during application of Strategy 2, the Concentrator, because of the Lexicon built by Strategy 1, will choose a Target Structure from a previous Focal Structure. Attached to the Focal Structure will be a relative-time-marker, which will become part of the Target Structure. Nodes which accept the tense-markers in the input string will be attaching processes which attempt to attach such relative-time-markers to the Parse.

Copula. There is little semantic weight in the copula. The BLOCKS world and most other semantic representations include explicit markers of some of its meanings, and the implicit analogues, as in (#COLOR :B1 #RED), are easily generated.

Articles. Like the copula, articles carry little semantic weight in the early stages of language use. For instance, if the current Focus contains a particular block, the commands "Pick up the block" and "Pick up a block" will both be responded to correctly if CLAP picks up the block under scrutiny. The phrases "a block with a pyramid on it" and "the block with a pyramid on it" present a similar choice, if there is a block with a pyramid on it.

In later stages, when the discourse pertains to non-Envircnmental referents, to concepts which have no perceivable referent, or to specific referents previously mentioned, articles will carry more import, as in

I wish I had a red ball.
Who is the Prime Minister?

In these examples, the article functions as a marker from which inferences can be drawn. In the first sentence, there could be many red balls; in the second, there is probably only one Prime Minister. Thus, in crder to be used correctly, the indefinite article must be associated with the existence of alternatives, the definite with the concept of uniqueness. These concepts are not available in the Focal Structure explicitly, but must be reccgnized by the Evaluator. Hence information about the quantification of concepts in the utterance must be passed by the Evaluator to the Parser-Modifier so that nodes can be built in the Parser which will interpret articles correctly. In general, any inference procedures and/or data which the Evaluator outputs will be passed in this way to the Parser-Modifier.

Third person: The third person inflection is largely redundant in comprehension. The structure built into the Parser will simply be a node following the verb node and will have no clear-cut meaning associated with it.

Auxiliaries: These segments are similarly redundant except for tense marking, so no special conceptual structures will be built for them.

It is clear then that some of the modulating morphemes are not crucial for comprehension at this stage. The fact remains, however, that the child and CLAP must gain control over them in order to communicate successfully. The methods of gaining this control of output are described in section 4.5.

It should be remembered that the processes of Strategy 1 continue into Strategy 2, with some augmentation. In Strategy 1 association weights were augmented when segment and concept co-occurred. In Strategy 2, CLAP can, in addition, increase an association weight when a concept participates in an input-output pair at a node of the Structure-Builder.

New ARTRANs are always built to share as much of the structure of the existing Parser as possible. For example, suppose the utterance that follows that of figure 4.6 is

B6 is on top of B7.

with the segment list

(B6 _is _on _top _of_ B7 .)

and Target Structure

(#SUPPORT :B7 :B6).

The new ARTRAN might look like the one shown in figure 4.7. However, if this were the only kind of generalization to go on, the Parser would become horrendously inefficient. To avoid this, Strategy 3 comes into effect.

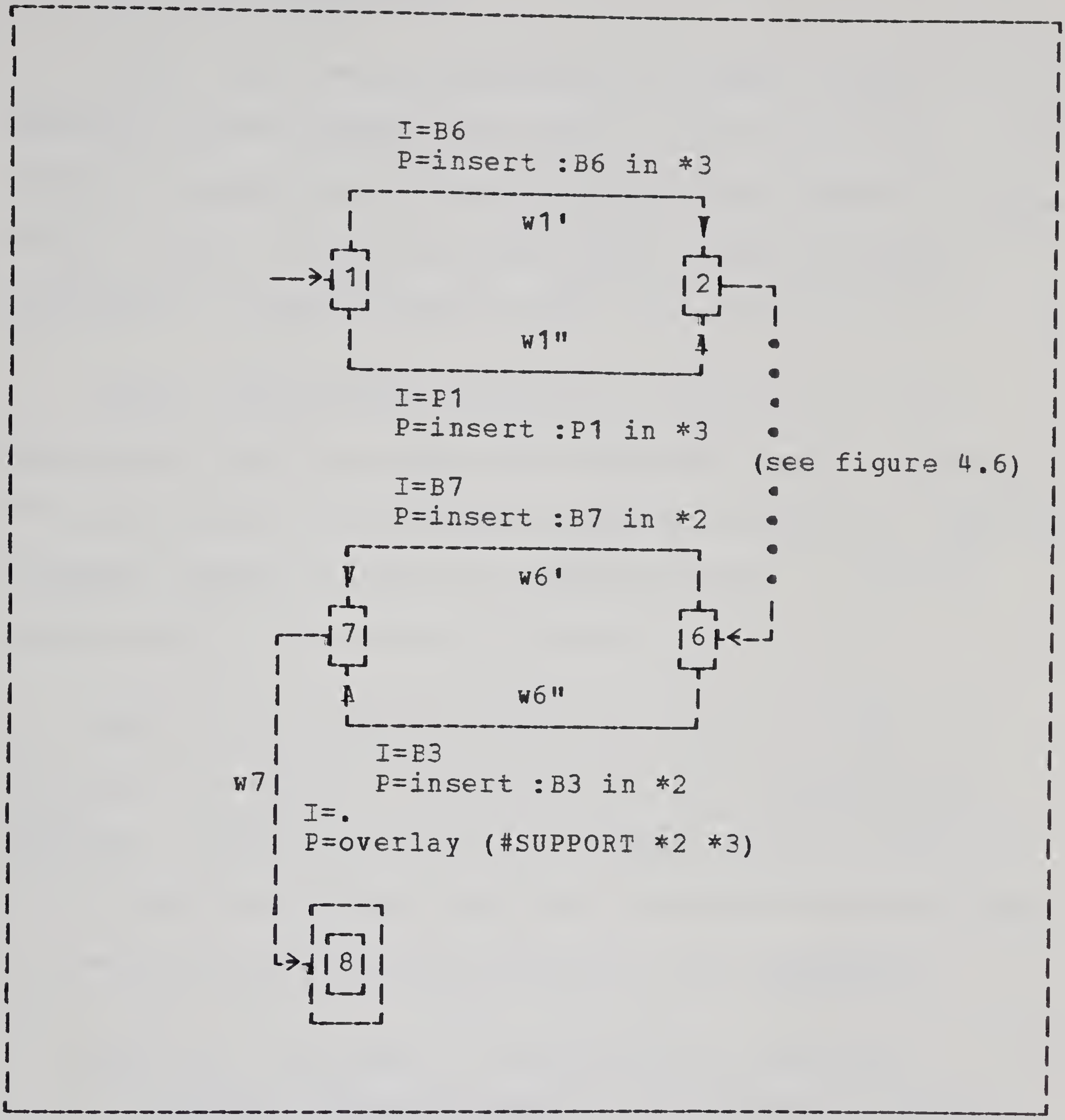


Figure 4.7: A modified ARTRAN (For an explanation of the notation, see Appendix 4.)

4.3.3 Strategy 3: structural generalization

This is a group of strategies which allows CLAP to make its Parser more efficient in terms of space, and gives it the ability to handle novel utterances without using the Environment. CLAP's process of generalization will have two

important characteristics.

First, the generalization will be based on the structure of the Parser, but will be directed by the semantic regularities of the processes which label the nodes of the Parser. There is, after all, no other external information on which CLAP can base its generalization.

Second, when nodes and arcs are combined into new, generalized nodes and arcs, the old nodes are not removed, but are allowed to age their way out of the Parser. This is necessary because there is no guarantee that a putative generalization is a correct structure.

One of the questions I will not answer is "When is Strategy 3 invoked?" because this is an implementation decision. Clearly it cannot be invoked before Strategy 2, but it could be invoked every time the Parser-Modifier acts, or every time the Responder-Modifier acts, *et cetera*.

There are two types of generalization possible:

1. *Semantic generalization*: reorganization of the Parser on the basis of the regularity of semantic characteristics of arc labels
2. *Syntactic generalization*: reorganization of the Parser on the basis of the topological similarity of substructures.

The sub-strategies of Strategy 3 combine these two criteria in various degrees.

Suppose A={a(i)} is a set of arcs, and each a(i) has associated with it a process P(i) and input I(i). Suppose also that there is a slot s common to all P(i)'s, and this slot is filled by each P(i) with the concept c(i). If these concepts share an attribute or a value, or have attributes which share an attribute or value, and so on, then a new arc b(i) can be created for each i.

Each b(i) is labelled with input I and process P. I is a branch to a procedure which attempts to find a segment which matches the attribute or value criterion described above. If it does, it returns c, the meaning of the segment, and P uses that meaning as each P(i) would have used the meaning of each I(i). Each b(i) is attached to the source and target nodes of a(i), in parallel with a(i), and is given a weight which is greater than a(i)'s.

As an example, suppose the Parser has the form in figure 4.7. The the arcs joining nodes 1 and 2 satisfy the above set of conditions: the slot *3 is common to the two processes, and the concepts :P1 and :B6 are related as follows:

(#IS :P1 #PYRAMID)

(#IS #PYRAMID <u>#THING</u>)

(#IS :B6 #BLOCK)

(#IS #BLOCK <u>#THING</u>).

Similarly the other pair of parallel arcs is related in at least one way:

(#IS :B3 <u>#BLOCK</u>)

(#IS :B7 <u>#BLOCK</u>).

In the first case, a new arc joining nodes 1 and 2 is created with weight w1"', and I, a call to a procedure which expects a segment such that its meaning c satisfies:

(#IS c x)

(#IS x #THING).

The new process attached to the arc will be "insert c in *3". Similarly for the other pair of parallel arcs.

This strategy in itself does not reduce the number of nodes and arcs, but it does generalize the processes on the arcs. The number of arcs will be reduced by the process of ageing, whereby arcs are deleted when their relative weights fall below a threshold. The weight on an arc b(i) will increase if CLAP experiences an utterance in which there is a word in such a position and with a meaning which can match the attribute associated with P. Thus utterances which would have used a(i) will use b(i) and its weight will increase. New utterances with new words which can also use b(i) will also increase its weight. Thus, in many cases, the original arcs {a(i)} will disappear and the new generalized arcs {b(i)} will remain.

I will call the above process Strategy 3.1. It builds structures which resemble feature-matching procedures, and is an example of semantic generalization.

Strategy 3.2 is a kind of syntactic generalization

which depends to some extent on 3.1. It is similar to what Harris <1972> called "Recursion". If two arcs in sequence have the same inputs and processes, then a new set of arcs is created to allow those inputs and processes to be carried out recursively. Clearly this will almost always be done with arcs that have been created by Strategy 3.1, since it is rare for the same lexical item to be repeated.

There are some major differences between Strategy 3.2 and Harris' Recursion strategy. In his case, the strategy is applied directly to the grammar. Here, as with all CLAP's strategies , it is applied to the Parser, and only affects the Responder indirectly. Harris' strategies have no means of self-correction. Once applied, the resulting productions remain in the grammar, right or wrong. In CLAP's case, a generalization which turns out never to be used, or which produces bad Parses, ages out of the Parser.

A second strategy for syntactic <u>and</u> semantic generalization is Strategy 3.3, which is related to Harris' "Grouping". Parts of the ARTRAN which are structurally congruent are used to create a new sub-ARTRAN which generalizes the structure-building operations of the original parts. This strategy can be invoked before and after Strategy 3.1, since it involves two kinds of generalizations. The first involves generalizing the position of structures in a Parse. For instance, in English, a locative phrase might take many positions in an

utterance, and hence, early in CLAP's life, structures for analyzing one such phrase might appear in many places in the Parser. Strategy 3.3 would create a single copy of this sub-ARTRAN and create an arc invoking it in parallel with each original sub-ARTRAN.

The other situation in which Strategy 3.3 will be invoked is when two sub-ARTRANs have the same shape but not the same operations on each arc. For instance, in French, a post-nominal modifier can be a prepositional phrase, an adjective, a relative clause, <u>et cetera</u>, but in general they are parsed in a similar way:

1. Accept the noun and put its concept into the Parse,
2. Look for a prepositional phrase,

<u>et cetera</u>, <u>mutatis mutandis</u>. Thus a new arc can be created which invokes a sub-ARTRAN analogous to:

1. Accept a noun and put its concept into the Parse,
2. Look for a $\left\{\begin{array}{l}\text{prep. phrase}\\ \text{adjective}\\ \text{rel. clause}\end{array}\right\}$

The three sub-strategies outlined here may not turn out to be sufficient for all generalization necessary, but given the ARTRAN framework it should not be difficult to postulate, and experiment with, various heuristics for building new ARTRANs from old.

4.3.4 Strategy 4: Parser versus Environment

In its early, naive stages CLAP has but one source of information about reality: the Environment. As its grasp of language increases, however, it gains a second, sometimes competing, source. For CLAP, language competes with reality in the following situations:

1. The Human lies to CLAP.
2. The Human utters a sentence or series of sentences which are concerned with a hypothetical world.
3. The Human utters a negative sentence which is true of the Environment.
4. The Human asks a question which resembles a statement.
5. The Human utters a command to produce a situation which does not now obtain.

The first case can be easily disposed of - if the Human does this he's asking for trouble! CLAP will be able to accept a number of counterfactuals in its early stages without being too confused, but if their number is too great it will have just as much trouble as a human child would have if most of the things he heard contradicted his experience. The other cases are more interesting, and require a major Strategy to handle them.

Initially, cases 2-5, like case 1, will add noise to CLAP's experience, and will not contribute a great deal to

its Parser, other than helping build some lexical entries. As the weights in the Parser become greater, and CLAP begins to put credence in its Parses comparable to that which it puts in its perception of reality, it becomes important for CLAP to assign interpretations to what appear to be counterfactuals; that is, to Parses which appear to conflict with the present state of the Environment.

It is unlikely that any of the correct interpretations of the kinds of utterances described above are innate in the child. It is possible that some are learned through non-linguistic means. Commands can be intimated by gesture, but questions are almost all dependent on language for communication. It is difficult to prove, but I conjecture that that negations of propositions cannot exist as concepts without the ability to express them in language. For instance, it seems unlikely that, without any knowledge of language, one would formulate a concept analogous to (#NOT (#COLOR #ELEPHANT #PINK)); the concept would always be of the form (#COLOR #ELEPHANT #GREY). The two other categories of negation, absence and non-existence, may be another matter. It is likely that one can formulate (#ABSENT :MOMMY), or (#ALLGONE #MILK) without language (perhaps they would both be #ALLGONE for the child). Lacking any sound theories as to how these counterfactual concepts arise, I propose that CLAP's learning strategy should be pre-programmed in the following manner.

When CLAP arrives at a (possibly incomplete) Parse of an utterance, and that Parse, or part of it, contradicts the Semantic Base, the Evaluator has the following choices:

1. Build a hypothetical world model into which the Parse will fit.
2. Negate the Parse.
3. Mark the Parse as a goal and hand it to the Action-Taker.
4. Extract the true state of affairs from the Semantic Base and pass it on to the Action-Taker for output.
5. If the Parse does not contradict the Envircnment, adjust the Semantic Base to agree with the Parse.

Assuming the Parser has no reason to make any particular choice, the Evaluator will choose one according to some heuristic, and mark the Parse appropriately. The Parser-Modifier will then treat such marking much as it treats any cther concept. Unknown segments of the utterance, like negative inflections, question markers, and conditional markers are associated (nct necessarily correctly), through the Parse, with whichever type of interpretation the Evaluator chooses.

As with all the other strategies, the structures built by this one will be modified as subsequent utterances are given other interpretations. Eventually segments will be associated, through their position in the ARTRAN which is

the Parser, with the correct interpretation of counterfactuals, negatives, commands, and questions.

4.3.5 Strategy 5: using discourse

Strategy 4 had to resolve conflicts between Parses and the Semantic Base; Strategy 5 uses Parses tc supplement or replace the Focus. There are many discourse phenomena which CLAP must learn to deal with; among them are:

1. anaphora,
2. implied scope of quantification,
3. hypothetical wcrlds,
4. referential ambiguity.

The use of previous utterances in analyzing the current one must wait until CIAP's Parses have enough weight, since it would be foolish to allow one tentative Parse to guide a second tentative Parse. The weights given to the Parses will ensure this, since they will initially be lower than the weights on the Foci. In order to make past Parses available to the Parser and Evaluator, they are stored, along with past Foci, in Short-Term Memory. Here they are aged out according to some memory-management scheme.

While systems exist which can handle the linguistic features learned by the first four Strategies, not much is known about discourse phenomena. Part of the reason for this lacuna is the concentration of linguists on the

sentence as the highest interesting syntactic category. For this reason, it seems desirable to wait until the results of the initial Strategies are seen, or until somewhat more is known about the underlying structure of extended discourse, before attempting to specify the nature of Strategy 5.

4.4 Learning to speak

Learning to understand a language is not all there is to acquiring that language; CLAP must learn to respond. Section 4.2 alluded to but did not describe the Responder and the Responder-Modifier.

4.4.1 The Responder

This component accepts as input the Intention constructed by the Action-Taker. Its output is an utterance - an ordered set of lexical items representing the Intention. The Responder may be described as an ARTRAN whose arcs are labelled with:

1. a pointer to an element of the Semantic Base, or
2. a pointer to a sub-ARTRAN which returns an element of the Semantic Base, or
3. a pointer to a lexical item or items (possibly null).

The first two may be considered as inputs to the arc, the third is the output. The Responder thus traverses the Intention and constructs an utterance from left-to-right. The detailed structure of the Responder will become more evident in the next section.

4.4.2 The Responder-Modifier

This component applies a uniform strategy to the conversion of Parser structures into Responder structures, in contrast to the several temporally interdependent Strategies of Section 4.3. Its inputs are the Lexicon and the Parser, and it affects the Responder and the Semantic Base.

In examining the Lexicon, the Responder-Modifier attaches to each concept in the Semantic Base those lexical items which have that concept as a clear-cut meaning. They are attached in a way similar to that in which any attribute is attached to a concept, so that they could conceivably be used by the cognitive component if that component were able to learn. When a lexical segment is attached, it is given a weight related to the weight on the segment→concept link. Each time the Responder-Modifier updates these links in the Semantic Base, the concept→segment weight $w(c,s)$ is changed in a way dependent on the current segment→concept weight $w(s,c)$. If $w(s,c) < w(c,s)$ or if $w(s,c)$ no longer exists,

then w(c,s) is reduced. If w(s,c) > w(c,s) then w(c,s) is augmented. This process maintains a balance between segment→concept associations and concept→segment associations. In this way meanings which are initially incorrectly learned are eventually discarded.

The operation of the Responder-Modifier on the Lexicon is fairly straightforward; its operation on the Parser is less so. Here it must examine the inputs and processes which label the arcs of the Parser and translate them into semantic conditions and graphemic outputs labelling arcs on the Responder-Modifier.

An input I(p) on a Parser arc a(p) may be:

1. a segment
2. a branch to a sub-ARTRAN
3. a set of conditions on the attributes of the meaning of the current input segment.

The process on an arc is a semantic structure or frame. The input I(r) on an arc a(r) of the Responder must be a pattern-matching routine which attempts to find the structure or frame in the Intention. The pattern-matcher returns either failure, success, or a subsequence of segments. The output is a segment or a sequence of segments. The mapping of Parser arcs to Responder arcs is as follows.

If I(p) is a segment, then I(r) is a pattern-match, and O, the output of the Responder arc, is the same as I(p). If

I(p) is a branch, then I(r) is a branch to a sub-ARTRAN whose arcs are constructed according to the arcs in the Parser sub-ARTRAN. The output is null, since arcs of the Responder sub-ARTRAN will output all the segments associated with the invoking arc.

The third possibility is that I(p) is a set of conditions on the attributes of the meaning of the segment. In this case, I(r) is a pattern-match, as before, but it returns a segment which is determined by the criteria in I(p) and the concepts in the structure or frame in P. The pattern-matcher checks the concepts in the Intention for a concept which matches the criteria. It returns the segment which has been associated with the concept by the Responder-Modifier's examination of the Lexicon.

Creation of arcs in the Responder will take place only for Parser arcs which have weights above a threshold. Thus comprehension will almost always precede and exceed production.

Clearly the details of the transformations performed by the Responder-Modifier are missing, as are the algorithms for traversing the Parser and selecting arcs. As we shall point out below, however, this is such an unexplored problem that such solutions are beyond the scope of this work. What I have sketched is a framework on which an experimental implementation could be based.

4.5 Possible implementations

Here I will attempt to point out existing systems, methods, algorithms, and representations which might be used for each component of CLAP. Some of them have already been mentioned in preceding sections of this chapter.

4.5.1 The Environment and Perceiver

As implied by the preceding sections, a CRT display of a line-drawing representing a 3-dimensional scene is, in my opinion, sufficiently rich to induce many of the processes of language acquisition which occur in humans. Winograd's <1971> BLOCKS world is such an Environment, and extensions of it are easily imaginable, limited mostly by the efficiency of transformation algorithms and hidden-line processing.

As the Environment becomes more complex, however, a question must be answered about the Perceiver. Are all relations between objects explicitly represented in the Semantic Base at all times, or are they generated as part of the Focal Structure? In the BLOCKS world, some relationships are represented explicitly unless they can be deduced from transitivity; an example is #SUPPORT. Other relations are produced in the process of goal-seeking, like "bigger than". Why the difference? The question is important to CLAP because if association is to take place,

concepts must be explicitly represented, or some radically different mechanism for initial vocabulary-acquisition must be postulated. Without explicit representation of a concept like #LEFT-OF, at least in the Focal Structure, CLAP would have to start generating all the relations it could think of in order to find something with which it could associate the word "left". Hence I suggest that the Perceiver should generate all reasonable inferences when it creates the Focal Structure.

4.5.2 The cognitive components

While the number of implemented semantic representations is very large, few have been integrated with a representation of an Environment. The exceptions are the BLOCKS world <Winograd 1971> and ENGROB <Coles 1968>. In implementing CLAP, it is desirable to use a Semantic Base which has proved its compatibility with the Environment which is the essential interface between CLAP and the Human. The BLOCKS representation is by no means the ideal Semantic Base, but it does satisfy the latter requirement. It also has the advantage of a uniform representation of actions, inferences, events, *et cetera* which makes it amenable to the use of uniform learning procedures in the Parser-Modifier. It remains to be seen how far the representation of the particular BLOCKS world can be generalized to more complex Environments.

Winograd's "event list" is the germ of CLAP's Short-Term Memory. However, there is really no extant example of this component, since it must contain not only representations of events in the Environment but Parses, responses, Focuses, and Human non-linguistic inputs. These might be kept on a stack, each with a weight dependent on the origin of the event and its time in Short-Term Memory. The weights would guide the search carried out by the Parser and Parser-Modifier. This component offers opportunity for research into the modelling of short- and medium-term memory in humans and its effect on the analysis of utterances.

4.5.3 The Parser

While the parsers of Winograd <1971>, Schank *et al.* <1973>, and Wilks <1973> are fairly successful, the primitives in their method of representation lack the simplicity necessary for an adaptive system. For this reason, Woods' <1969> Augmented Recursive Transition Networks (ARTRANs) are proposed as a representation which allows a uniform adaptive procedure to be applied to the Parser. Since Winograd <1972> has already adapted Woods' structures to the production of SHRDLU's semantic structures, there is reason to believe that they would be adequate for CLAP's Parser.

4.5.4 The Parser-Modifier

Olivier's <1968> segmentation-learning program is the prototype for CLAP's Strategy 1, augmented with the ability to use meaning as a clue to segmentation. The structure outlined by Cercone <1975> is the best candidate for representing the Lexicon, since for one thing it lends itself easily to arbitrary division and combination of lexical items. The word-concept association scheme proposed is similar to Harris' <1972> procedure, which was successful in the extremely limited environment and artificial training regime of his robot. The method of weighting the arcs of the Parser has a remote similarity to Jordan's <1972> scheme for augmenting weights on her association net. Apart from these implementations, the strategies outlined in section 4.3 are experimental and speculative.

4.5.5 The Responder

Simmons and Slocum <1972> have described a discourse-generator whose input is a semantic net and whose output is a sentence or sequence of sentences. It is represented as an ARTRAN, and is the inspiration for the Responder described here. Clearly, however, what is at stake is not whether an ARTRAN can generate sentences (it obviously can), but whether CLAP can generate the ARTRAN.

4.5.6 The Responder-Modifier

As mentioned above, there is apparently no work on the transformation of a natural-language parser into a grammar or sentence-generator. Formal grammarians have concentrated solely on the reverse process. Hence, again, there are no existing implementations to draw from and the suggested transformation is purely speculative.

4.6 Linguistic implications

The child neither makes nor tests syntactic hypotheses; he attempts to create incrementally more complex rules for deriving some kind of sense from the utterances he hears. Generalization first takes place when the semantic conditions for a rule are such that they are met by new utterances as well as the original. It later takes place when similar rules are conflated. Association is an important element of language-acquisition strategies, but an explanation of acquisition must combine cognitive association with complex structures to be successful.

The phenomena of child language are a result of production rules which are built from comprehension rules already learned. Hence a more representative but certainly less accessible object of study is the set of utterances and utterance-segments that the child <u>understands</u> at an instant in time. In fact, a general implication of the ideas

presented here is that the proper study of language is the relationship between form and meaning, not among forms.

4.7 Summary

The learning schedule which CLAP undergoes is a realistic flexible sequence of linguistic and non-linguistic inputs and outputs. CLAP applies a sequence of strategies to these inputs, each strategy being initiated only when the previous strategy has built up sufficient structures. The earlier strategies have been described in much more detail than the latter ones, since even less is known about them. A method for building production mechanisms from comprehension procedures has been postulated. I have indicated some of the existing techniques which could be put together in an experimental version of CLAP. The last section briefly stated some testable linguistic implications of the model proposed here.

5

VAS: A First Step

5.1 Introduction

VAS (Vocabulary Acquisition System) is an attempt to use the framework of an existing language-processing system to accomplish one of the first tasks of a language acquisition system: the attaching of meanings to words. The Envircnment is simple, and VAS has low cognitive ability, but the linguistic input is unrestricted in form.

5.1.1 The Environment

As figure 5.1 shows, VAS' world is superficially the same as the BLOCKS world of Winograd's <1971> robot SHRDLU. schwa Unlike SHRDLU, however, VAS cannot manipulate objects; they change position by fiat. The Environment is, of course, much simpler than that of a human child, for all the objects in it are inanimate, uniformly coloured polyhedra,

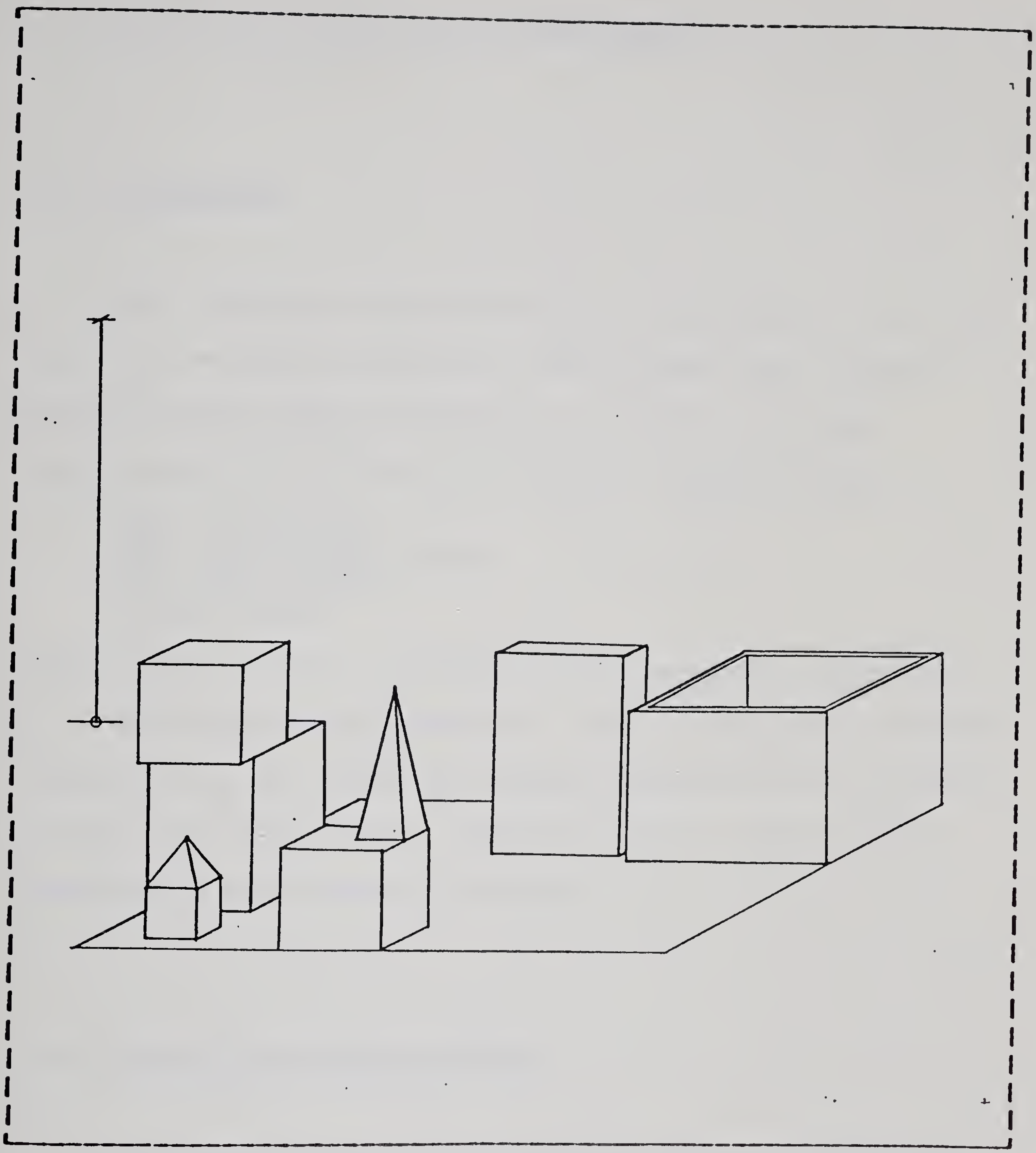


Figure 5.1: VAS' world

lacking textural differences and changing position instaneously rather than with continuous motion. There are no shadows or other variations in lighting. Objects have no resilience, temperature, or weight for VAS. VAS is not a robot that can move around the Environment, like a child, so

it always has the same view of the scene.

5.1.2 Cognition

VAS' internal representation of the scene is a set of one- and two-place predicates which encode such things as attributes and class membership of objects and other predicates, and positions of objects. Examples are:

```
(#IS :B1 #BLOCK)
(#AT :B2 (100 200 200))
(#IS #RED #COLOR)
(#MANIP :BOX)
```

(See Appendix 1 for a complete list.) Geometric properties of objects are not represented. None of Winograd's PLANNER theorems used by SHRDLU for drawing inferences and problem-solving are used by VAS. There is also no capacity for remembering past states or events.

5.1.3 Linguistic representation

The object of VAS is to build a rudimentary Lexicon, the form of which is constrained by the form of Winograd's DICTIONARY. A word which occurs in an input sentence can be incorporated by placing two properties on its property list: the indicator WORD with value (WORD) and the indicator SMNTC with the value

```
(WORD ((MEANS ((<concept> <weight>)
```

(<concept> <weight>)...)))).

For SHRDLU, the first element in the value of SMNTC would encode the part of speech of the word, like NOUN, ADJ, etc., instead of WORD, but at this stage VAS does not differentiate words syntactically. The formation of the list of (<concept> <weight>) pairs is described in section 5.2.3.

It is clear that at this stage any linguistic representation which includes a lexicon could incorporate the simple list of (<concept> <weight>) pairs.

5.1.4 Linguistic input

Each sentence is input in upper case, with complete punctuation, including terminal punctuation marks. Each sentence is segmented into words automatically, using blanks and other punctuation as separators. Hence there is no learning of segmentation, as there would be for a child.

5.2 The learning schedule

Ideally, within the constraints of the model described so far, the learning schedule would consist of a set of pairs of the form (s,f), where s is a sentence and f is a point, called a focal point, in the 2-dimensional

representation of the BLOCKS world. The point f would determine the focal region of the scene on which VAS was to fix its attention, just as a parent or some other aspect of the child's environment draws its attention to a particular part of its surroundings <Neisser 1966>. However, the available time and graphics software support have not allowed the implementation of such a form of input. Instead, the initial "pre-attentive" <Neisser 1966> processing has been done by hand for VAS.

Thus the input consists of a set of pairs (s,f), where f is a list of the internal names of the objects which are noticeably contained within the focal region selected as connected with s. The manner in which f is constructed, given the sentence s, is described in section 5.2.2.

As VAS reads each (s,f) pair it creates and modifies weighted links between words in the sentence and elements of the focal list. This is the second stage in which time and resource limitations have necessitated a less than ideal design. Memory management should be such that if a word is not experienced often enough, it is deleted from the Lexicon. Similarly, if a word continues to build up associations which are uniformly distributed over VAS' concepts, without a clear-cut meaning emerging (as is the case with articles), then at this stage of the acquisition process, it should be made inactive. It should take no further part in association-building, but should await a

later stage at which either new concepts are added or syntax starts to be learned.

Such strategies were not implemented in VAS; instead, a list of words taken from the corpus is input initially, and it is this Lexicon into which VAS incorporates meanings as a result of its experience with (s,f) pairs.

Having experienced a set of (s,f) pairs, VAS can be asked to output the likeliest candidate for meanings of the words in the Lexicon, and can save its accumulated associations for further later learning sessions. These operations are described in section 5.3.

5.2.1 The corpus

Two corpora were used in the experiment with VAS. In each case, the same materials were used. A drawing was made of each of the scenes depicted in figures 1, 6, 7, 11, 12, 15, and 19 of Winograd's <1971> thesis. They were coloured, and labelled not with colour names like the originals, but with object names: B1, B2, P1, P2, BOX, etc. (See figure 5.1.)

For corpus 1, an adult university graduate was given each drawing in turn and asked to talk, in as natural a way as possible, about the scene depicted in the drawing. Her discourse was recorded and transcribed, preserving the order

of the sentences, intc machine-readable form.

For corpus 2, I concocted for each drawing sentences which reflected (a) only concepts which VAS knows, and (b) only facts which were new for that drawing. That is, in describing relaticnships between objects, I mentioned only those which had changed since the previous drawing.

5.2.2 The Foci

Each focal list, or focus, was constructed according to the following (highly subjective) rules:

1. Read a sentence from the corpus.
2. Decide where the attention is directed in the scene. There may be 0,1 or more places to which your attention is directed. If there are none, then the focus is NIL. If there is one or more, apply step 3 to each cf your focal regions.
3. Place the template[1] over the focal region. Add to the focus the internal name of each object which has a major portion of itself within the area outlined by the template.

---

[1] The template is a piece of cardboard with a hole in it which displays only a portion of the scene. It simulates a simple windowing technique which would be used if the BLOCKS world were displayed on a CRT.

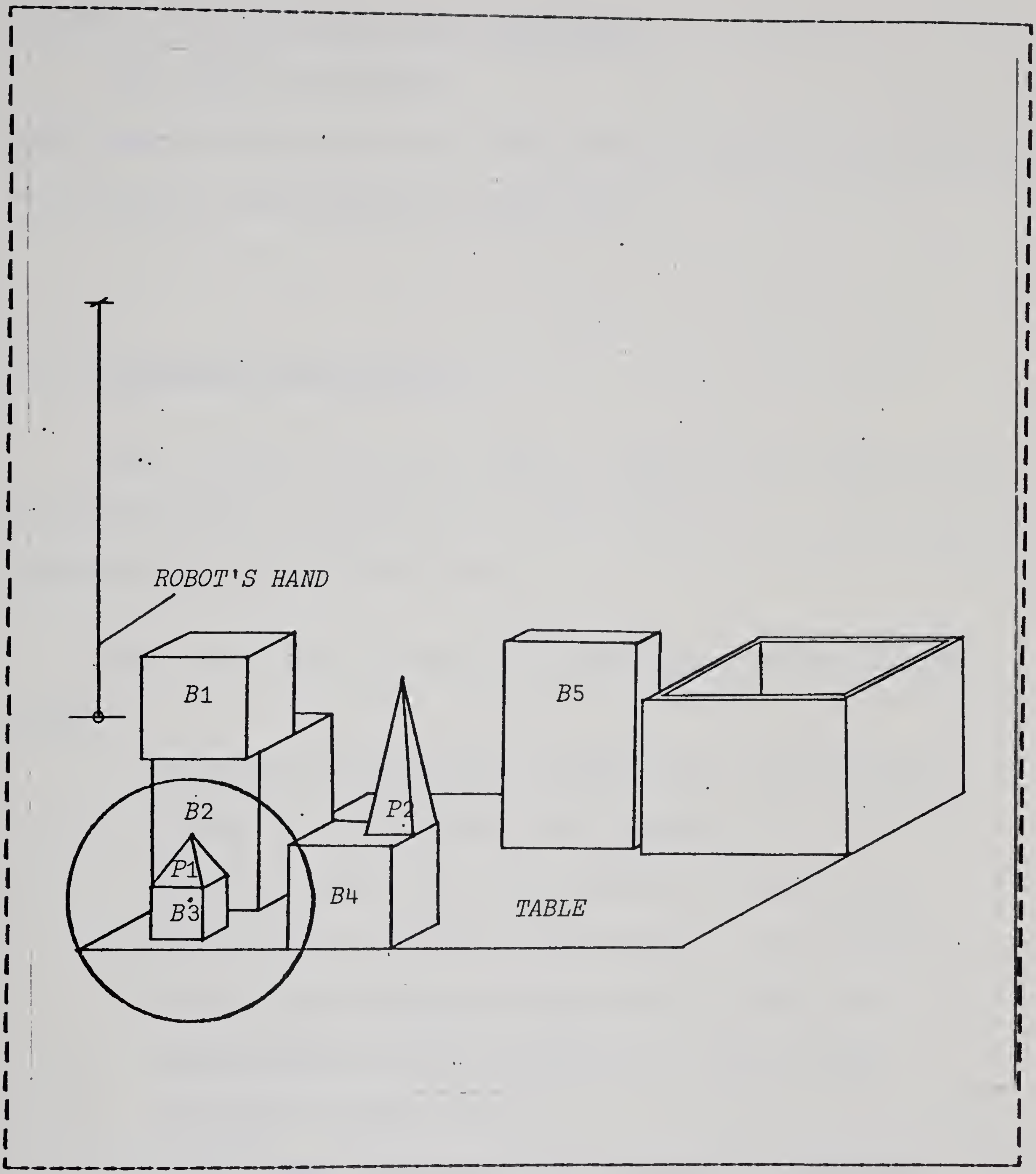


Figure 5.2: A focal region for the sentence "There is a houseshape made of a red cube, B3, topped by a green pyramid, P1, directly facing me to the left of the table."

For example, for the sentence

> There is a houseshape made of a red cube, B3, topped by a green pyramid, P1, directly facing me to the left of the table.

the focal region shown in figure 5.2 was chosen. The

objects in this region give the focus

(:B1 :B2 :B6 :TABLE).

(The names in the list are the internal names of P1, B3, B2, and TABLE. See figures 5.3 and 5.4.)

5.2.3 Building associations

When an (s,f) pair is input, a list of the words in s is constructed, stripping off all punctuation. The focus is expanded in the following way.

There are three classes of predicates in the BLOCKS world:

1. One-place attributive predicate. For example, (#MANIP x) attributes the property of manipulability to each concept in the list x.
2. Two-place attributive predicate. For example, (#IS x y) attributes membership in the set represented by the concept y to each of the concepts in the list x.
3. Relational predicate. For example, (#SUPPORT x y) means that the non-commutative and transitive relation of "supporting" exists between the object x and each object in the list y.

Each concept c in f is examined, and each class of predicate is processed as follows:

1. For each Class 1 predicate p in which c occurs, p is added to the focus.
2. For Class 2 predicate p in which c appears as a first argument, each concept in the second argument is added to the focus.
3. For each Class 3 predicate p in which c occurs as the first argument and one of the concepts in the second argument also appears in the focus, p is added to the focus.

Notice that predicates are concepts and may appear in the Focal List.

The above processes are applied in such a way that the focus is their closure. That is, if they were again applied to the focus, nothing would be added. Thus the expanded focus is a list of all the concepts which are relevant to the focal regions.

Having thus converted the (s,f) pair to a pair of lists, the process of associating words with concepts begins. Each word in the Lexicon has associated with it a (possibly null) set of concepts {c(i)} and weights {u(c(i),w)}. Each word or concept has associated with it a usage counter u(w) or u(c) respectively, whose value is the number of times the word or concept has appeared.

For each lexical entry w which is in s, the usage count is incremented. Then for each concept c(i) associated with w, u(c(i),w) is incremented if $c(i)\in f$. In addition, if $c\in f$

and there is no link between w and c, then a link is added to w with a weight of u(c,w)=1. The usage count of each concept in f is incremented by 1. The result of this process is equivalent to a matrix in which the ijth entry is is a count of the number of co-occurences of the ith word and the jth concept.

5.2.4 Deriving a meaning

After a number of (s,f) pairs have been input, each word in the Lexicon will have a set of concepts associated with it, a weight on each link with a concept, and a count of the usage of the word. Each concept will also have a usage count. The associations built up fall roughly into the following categories:

> Case 1: The word occurs in a large percentage of the sentences, and the associated concept occurs in a large percentage of the foci. For example, "the" occurs in most sentences, and the relative position #BOTTOM occurs in most foci. Hence a large association weight builds up between "the" and #BOTTOM. It is generally true that the words with the highest frequency of occurrence are unlikely to have a concrete referent or even a high-level conceptual referent like #COLOR or #SUPPORT. They are also not among the first words

uttered by the human child, and hence not among the first words understood. Similarly, concepts which are almost always present are less likely to be noticed by a child. If a high association value were used to assign a high-frequency concept as the meaning of the word, many (perhaps all) words would be assigned the most frequently occurring concept as their meaning.

Case 2: The word occurs in a small percentage of the sentences and the concept occurs in a large percentage of foci. As in 1, it is unlikely that the concept should be assigned as the meaning of the word, since it is the high frequency of the concept only that causes a high association weight.

Case 3: The word occurs in a large percentage of the sentences and the concept occurs in a small proportion of foci. As in 1, it is unlikely but possible that a meaning can be chosen from the high-level concepts of the BLOCKS world. However, it is also unlikely that the association measure is very high compared with other concepts associated with the word, since more frequent concepts will have had more chances to build up links with the word.

Case 4: The word occurs in a small proportion of the sentences and the concept in a small

proportion of the foci. This presents the most promising possibility for a meaning, since an association will be built up only if, on the few times that the word occurs, the concept also occurs.

One way to choose a meaning for a particular word w is to define a function F involving the following three quantities:

$u(c)$ the usage of an associated concept c

$u(w)$ the usage of w

$u(c,w)=u(w,c)$ the number of times c and w have co-occurred.

The function chosen here is

$$F(w,c) = u(c,w).(2-m.u(c)/u(w))$$

The choice of the positive coefficient m is outlined below.

F has the following properties:

1. If the ratios $u(c)/u(c,w)$ and $u(w)/u(c,w)$ are held constant, F is a linear increasing function of $u(c,w)$.
2. If $u(w)$ is constant, and $u(c,w)/u(c)$ is constant, then (a) F is a concave-down quadratic function of $u(c)$, whose maximum is at a point determined by m and $u(w)$, and (b) the value of $u(c)$ at which F is a maximum is

$$u(c) = u(w)/m$$

and the value of $\max\{F\} = u(c,w)$, so the value of $u(c)$ such that F is a maximum is an

increasing function of $u(w)$, and $\max\{F\}$ is an increasing function of $u(c,w)$.

Property 1 means that words with high associations will be ranked high. Property 2(a) means that concepts with very high or very low frequency of occurrence will be ranked low, if $u(c,w)/u(c)$ is constant, and the highest ranked concepts will have usage counts somewhere in the middle of the interval $(0,t)$, where t is the maximum possible usage count.

Property 2(b) is important because it mollifies 2(a) somewhat. If a concept has high usage, but the word also has a high usage, then that concept will <u>tend</u> to get a higher rank than it would have if the word usage were low.

For each word in the Lexicon the list of associated concepts is sorted by decreasing value of $F(w,c)$, and the first concept in the list is printed. The sorted list is retained, and would be useful in later parsing, when it would provide a plausibility ranking which would allow the Parser to choose the interpretations of words in decreasing order of plausibility.

The coefficient m used in the formula for $F(w,c)$ was at first assigned the arbitrary value of 1. By experiment, it was found that the highest score of correctly learned words was achieved with $m=.21$. Whether this value is optimum for other corpora or over time could be determined by further experiment.

5.3 Results

Since VAS has no component that models overt behaviour, I have used internal data structures for evaluating its success. The evaluation is subjective in the sense that criterion for determining whether a word is assigned a correct meaning is whether I think the meaning is correct. The words in the Lexicon were chosen from the corpora on the basis of their potential "learnability"; that is, on the basis of the existence of concepts with which they could be associated, and the existence of diverse encugh situations to build up distinctive associations. Words not potentially learnable (in my estimaticn) were eliminated to cut down on processing time, since VAS has no method for removing unlearnable words automatically. The processes of building associations and choosing a meaning is unaffected by this limitation of the Lexicon. (For further discussion see the last part of section 5.5.1.)

The results for each corpus are summarized in figures 5.3 and 5.4.

There are 16 frequently-occurring words in corpus 1 which have directly ccrresponding concepts, and which could conceivably be learned by VAS by association. That is, they occur in diverse enough situations that there is (subjectively) a chance of discriminating among the many concepts with which they co-ccur. These wcrds were placed in VAS' Lexicon, and VAS was given the list of 219 (s,f)

Word	VAS' meaning	Correct meaning	Reason for error[1]
BLUE	:B8	#BLUE	UCC
PYRAMID	#GREEN	#PYRAMID	MEW
P3	:B4	:B4	
BOX	:BOX	:BOX or #BOX	
HAND	:HAND	:HAND or #HAND	
RED	#RED	#RED	
CUBE	:B1	#BLOCK	ECU
B3	:B1	:B1	
GREEN	:B3	#GREEN	ECU
P1	:B2	:B2	
B2	:B6	:B6	
B1	:B7	:B7	
BLOCK	:BOX	#BLOCK	ECU
B5	:BOX	:B8	UCC
B4	:B3	:B3	
P2	:B3	:B5	UCC

Figure 5.3: Results using corpus 1
[1]See text.

pairs corresponding to the various configurations of the BLOCKS world, an average of 31 utterances per scene. These utterances, as Appendix 2 shows, are rambling, semantically noisy remarks on the various scenes, hardly suited to the task of training anything or anyone in the meaning of the words in the Lexicon. Of the 16 words, 9 were learned successfully; that is, they were associated with concepts which, to me, are their meanings within VAS' representation of the BLOCKS world.

VAS was more successful with corpus 2. Of the 24 learnable words, 18 were learned correctly. There were 39

[illegible]

Word	VAS' meaning	Correct meaning	Reason for error[1]
BLUE	#BLUE	#BLUE	
PYRAMID	#PYRAMID	#PYRAMID	
P3	:B4	:B4	
BOX	:BOX	:BOX or #BOX	
HAND	:HAND	:HAND or #HAND	
RED	#RED	#RED	
B3	:B1	:B1	
GREEN	#GREEN	#GREEN	
P1	:B2	:B2	
B2	:B6	:B6	
B1	:B7	:B7	
BLOCK	#BLOCK	#BLOCK	
B5	:B8	:B8	
B4	:B3	:B3	
P2	:B5	:B5	
WHITE	:BOX	#WHITE	UCC
BLACK	:TABLE	#BLACK	UCC
TABLE	:TABLE	:TABLE or #TABLE	
SUPPORTS	#SUPPORT	#SUPPORT	
LEFT-HAND	:B2	#LEFT	ECU
FRONT	:TABLE	#FRONT	ECU
BACK	#CENTRE	#BACK	ECU
CENTRE	:B5	#CENTRE	ECU or UCC
RIGHT-HAND	:BOX	#RIGHT	ECU

Figure 5.4: Results using corpus 2
[1]See text.

(s,f) pairs, with 19 in the first scene and an average of 3 per scene in the last 6 scenes.

There are three identifiable situations in which VAS does not learn the correct meaning of a word:

1. The scenes in VAS' experience are such that two concepts c1 and c2 always co-occur. In this case, differences in rounding off

weights, or chance ordering of the concepts will decide which of c1 and c2 is chosen. This case is called Uniform Concept Co-occurrence (UCC), and parallels a common error in children.

2. The corpus contains utterances of the following type: the utterance contains a lexical item whose meaning does not occur in the focus; that is, the utterances contain Misleading Extraneous Words (MEWs).
3. The correct meaning of a word w is a concept c whose high usage u(c) lowers the value of F(w,c) to a point where c is not chosen as the meaning of w. This case is called Excessive Concept Usage (ECU).

Figures 5.4 and 5.5 show which of these categories VAS' mistakes fall into.

UCC might be solved by introducing a "salience" factor which would serve to heighten certain concepts on the basis of their perceptual prominence. MEW is a fault of the corpus, and can only be corrected by conditioning the corpus to be less noisy or by postponing the learning of problem words to later strategies. ECU might be avoided by changing the weight-incrementing scheme to increment u(c,w) as many times as a concept occurs in the focus, thus increasing F(c,w).

The results of these limited experiments should be encouraging. The indication is that initial vocabulary acquisition by association can take place under two extremes of training. Corpus 1 provides vague, rambling, noisy input, while corpus 2 has low noise, and is not misleading. Acquisition is poorer with corpus 1, but VAS still manages some learning. The reasons for incorrect learning are identifiable, and there are possible ways of correcting them.

5.4 VAS in CLAP

There is no single component of CLAP analogous to VAS. VAS is a pilot program; it attempts to demonstrate that Strategy 1 (see section 4.3.1) of CLAP has at least a chance of working. Remember that this Strategy was concerned with two things: learning to find the meaningful segments (morphemes) in an utterance, and building up knowledge of the meaning of thes segments. VAS does not address the first goal; as for the second, VAS' goal is related.

Characteristics of VAS which connect it with the second goal of Strategy 1 are:

1. VAS (trivially) creates a Segment List.
2. VAS creates a Focal Region and a Focal Structure.
3. VAS builds weighted associations between

segments and concepts.

There are, however, major differences:

1. The Segment List consists of words, not putative morphemes.
2. The characteristics of the Focal Region are simplistic, not reflecting any hypotheses about scene perception.
3. The Focal Structure is a simple list of concepts, not a semantic structure.
4. The association weights are simple frequency counts, ignoring such factors as relative and inherent perceptual salience of concepts and words.
5. VAS' Lexicon is pre-specified because the corpus is known, hence the problem of knowing which words VAS should consider as learned never arises.
6. The concepts available to VAS are a small subset of those available to SHRDLU <Winograd 1971>, so the set of potentially learnable words is very small.

All these shortcomings can be translated into proposals for extending VAS.

5.5 Future work

VAS is a crude first step on the road to CLAP. There

are many possibilities for experiment with VAS, improving on it, and implementing alternative schemes.

5.5.1 Extensions

Two extensions have already been mentioned in the description of VAS. The first is the introduction of a memory-management scheme which would replace the manual selection of a Lexicon. Words could change their status depending on the pattern of their usage. This would involve keeping track of, for instance, the time since a word was last used. Words with rare or extremely frequent usage would be dropped from the active Lexicon. Similar strategies could be applied to concepts, so that concepts or whole classes of concepts would no longer be associated with words, reducing some of the processing time for (s,f) pairs.

In order to model human acquisition, some kind of "salience factor" could be assigned to each concept, based on human perception. This factor could be assigned to a concept either globally or during the processing of a focus, or both. It could also be a function of the usage of that concept, since an instance of a concept appearing for the first time is more likely to be salient than one seen often. It would be advisable to investigate contributions in the study of perception before introducing such a factor.

The other extension mentioned previously is the

implementation of CRT display of the Environment, so that foci could simply be pointed to. This would allow experimentation with various sizes of Focal Region, and various algorithms for constructing the Focal List.

There is also obviously much room for experiment with various non-linear ways of incrementing usage counts, meaning-selection functions, and so on. There may be results from psychology which could be used to construct the meaning-selection function.

Another obvious extension is to include all SHRDLU's <Winograd 1971> semantic capabilities and its ability to manipulate objects in the learning process. This would allow the learning of verbs, and perhaps prepositions and adverbs.

Perhaps the most important problem to be examined is that specifying criteria by which VAS can judge whether indeed a meaning can be confidently attached to a word. I have avoided this problem by pre-specifying the Lexicon, but if VAS is to be a complete pilot program for Strategy 1 of CLAP, it has to find some way of deciding, on the basis of u(w), u(c), and u(c,w), what degree of confidence it can have in the best candidate for a meaning. One way of establishing criteria would be to examine the associations built for several corpora to see what characteristics the values of u(c), u(w), and u(c,w) have for those words whose meanings <u>are</u> learned correctly. Hopefully a mathematical

criterion can be specified such that relative weights of association will be sufficient to choose those words whose meanings have been learned.

5.5.2 Experiments

There is an interesting experiment that could be carried out with VAS, as follows. Construct a model of the BLOCKS world. Present the model to a mother and a child who is at the stage just before uttering his first word (choosing this stage is, of course, somewhat problematical!) Have the mother talk to the child about the BLOCKS world, in its various configurations, and record the discourse and gestures of the mother on videotape or film. Use her conversation and gestures to construct (s,f) pairs for VAS, and see how well VAS learns with such input relative to the other corpora. There are two interesting independent variables in the experiment: (a) the correctness of VAS as a model of the child, and (b) the skill of the mother as a teacher. If VAS is a good model of the child's acquisition method, and if the mother's conversation is better suited to the child's learning methods than the narrative speech of the other corpora, then either more words will be learned from the mother's corpus, or words will be learned sooner.

Clearly experimentation with other corpora and other languages would prove interesting. One could attempt to

establish whether the optimum value of the parameter m is independent of language and corpus and whether there is variation in the size or type of corpus necessary to attain a given level of proficiency. Then perhaps one could try to determine what the parameter means!

5.5.3 Alternatives

The linguistic input to VAS is automatically segmented into words using blanks and other punctuation. Thus there are no mechanisms available for learning the meanings of morphemes and uninflected forms, as a child does. "Block" and "blocks" are learned separately by VAS, whereas if the stem were learned, progress should be far quicker.

The alternative is to precede the word-association stage with a segmentation learning stage. Olivier <1968> has written a program which learns to segment English text from which blanks and punctuation have been removed. This program could be used to learn to segment ordinary English text containing blanks and punctuation, and I conjecture that it would learn to segment at morpheme boundaries. The learning would not be perfect, of course, since English, (and other) spelling tends to obscure relationships sometimes ("contrapuntal" and "counterpoint"; "solve" and "solution"), and to mislead ("the", "there", "then"; "but", "butter"; "be", "beer", "best"). But in the early stages

this kind of confusion would be unimportant. As segmentation was learned, VAS' association procedure would continue, and learning should be both more efficient and wider in scope.

It is clear that the procedures VAS uses for creating the focal lists, building associations, and deriving a meaning could be applied to other systems besides Winograd's <1971>. However, the only other known system with an Environment incorporated is Coles' <1969> ENGROB. It is probable that methods similar to VAS' can be applied to his system.

There are doubtless many other possible schemes for incrementing association weights and choosing meanings. They could easily be inserted into VAS' routines and experimented with.

6

Conclusion

Linguists have supplied Artificial Intelligence with data and hypotheses relevant to language acquisition. Unfortunately, their data consist mostly of observations of utterances over time, sometimes augmented with descriptions of the conditions under which the utterances were produced. Few data have been offered to give an idea of the progression of comprehension in humans. Hypotheses, likewise, have been concerned with postulating the order in which rules are internalized by the child. Few hypotheses have been made about processes, and those that have are generally clumsy and ill-defined. Schwarcz <1967> is an exception. He has offered a fairly comprehensive, clear, and coherent description of a five-stage natural language acquisition system. I do not agree with all his assumptions; notably I don't think a formal language is necessary for semantic input, except as noted in section 4.2. However, Schwarcz' description is still a valuable paradigm.

I have tried to describe the range of characteristics possible in a practicable acquisition system, and the criteria by which they could be judged. Based on these criteria, the few existing acquisition programs are not very satisfactory. The best (most successful and realistic) of these, Harris' <1972> robot, applies (apparently independently) some of Schwarcz' suggestions. He divides the acquisition process into stages similar to some of Schwarcz', and inputs semantic information in a formal language. But, as Chapter 3 points out, he still falls far short of a realistic acquisition system.

CLAP is a hypothetical system which might display some of the characteristics of children's development that linguists have reported. Some of Schwarcz' ideas are represented in CLAP, but its most important departure from other systems lies in the hypothesis that comprehension precedes production, and that structures built for parsing become the stuff of which production mechanisms are built. Many of the tools necessary to build CLAP exist in other systems, and it is my conjecture that at least the first three Strategies of the system described here could be implemented now, though this is certainly no trivial task. The remaining Strategies need further detailed specification and experimentation before they can be programmed.

Part of my confidence in CLAP's plausibility stems from VAS' success in learning vocabulary by association. There

are many other possible experiments with VAS which could strengthen or weaken this confidence. It is my hope that these experiments will be performed, and the extensions and alternatives described herein will be explored. It is my further hope that CLAP, or some perturbation of it, will eventually be implemented, to enable humans to understand language acquisition better, and computers to understand language better.

References

Agin, Gerald J., and Binford, Thomas O. <1973> "Computer description of curved objects" *Proceedings of the third IJCAI* pp.629-640

Anderson, John R., and Gordon H. Bower <1973> *Human associative memory* V.H. Winston and Sons, Washington D.C.

Bloom, L. <1970> *Language development: Form and function in emerging grammars* MIT Press, Cambridge, Massachussetts

Bowerman, Melissa <1973> *Early syntactic development* Cambridge University Press

Braine, Martin S. <1971> "Two models of the internalization of grammars" in Slobin, Dan I. (ed.) *The ontogenesis of grammar* The Academic Press, New York, pp.153-186

Brown, Roger <1973> *A first language: The early stages* Harvard University Press, Cambridge, Massachussetts

Cercone, Nicholas J. <1975> *State based conceptual representation for natural language understanding*

Ph.D. Thesis, University of Alberta, Edmonton, Alberta, Canada (in preparation)

Chomsky, Noam <1957> Syntactic structures Mouton and Company, The Hague

--- <1964> "Current issues in linguistic theory" in Fodor, J.A., J.J. Katz (eds.) The structure of language Prentice-Hall, pp.50-113

--- <1965> Aspects of the theory of syntax MIT Press, Cambridge, Massachussetts

--- <1968> Language and mind Harcourt, Brace, and Jovanovich, New York

--- <1970> "Remarks on nominalization" in Jacobs, R., P. Rosenbaum, (eds.) Readings in English Transformational grammar Ginn and Company, pp.184-221

Church, Joseph <1961> Language and the discovery of reality Random House, New York

Coles, L.S. <1968> "An on-line question-answering system with natural language and pictorial input" Proceedings of the 23rd National conference of the ACM p.157

Dale, Philip S. <1972> Language development: Structure and function The Dryden Press Inc., Hinsdale, Illinois

Derwing, Bruce L. <1973> Transformational grammar as a theory of language acquisition Cambridge University Press

Earley, J.C. <1966> Generating a recognizer for a BNF Grammar Computing Centre paper, Carnegie-Mellon University, Pittsburgh, Pennsylvania

Ervin-Tripp, Susan <1971> "An overview of theories of grammatical development" in Slobin, Dan I. (ed.) *The ontogenesis of grammar* The Academic Press, New York, pp.189-212

Feldman, Jerome A., and Robert F. Sproule <1971> "System support for the Stanford hand-eye system" *Proceedings of the second IJCAI* pp.183-189

Feldman, Jerome A., et al. <1971> "The use of vision and manipulation to solve the 'instant insanity' puzzle" *Proceedings of the second IJCAI* pp.359-364

Fillmore, Charles J. <1968> "The case for case" in Bach, Emmon, Robert T. Harms (eds.) *Universals in linguistic theory* Holt, Rinehart, Winston, Inc., pp.1-88

Fillmore, Charles <1969> "Toward a modern theory of case" in Reibel, David A., Sanford Schane (eds.) *Modern studies in English* Prentice-Hall Inc., Englewood Cliffs, New Jersey, pp.361-375

Green, L.E.S., E.C. Berkeley, C. Gotlieb <1959> "Conversation with a computer" *Computers and Automation* 8:10, pp.9-11

Halliday, M.A.K. <1973> *Explorations in the functions of language* Edward Arnold, London

Harris, Larry R. <1972> *A model for adaptive problem solving applied to natural language acquisition* Cornell University, Ithaca, New York

Jacobs, Roderick A., Peter S. Rosenbaum <1968> *English Transformational Grammar* Blaisdale, Waltham,

Massachussetts

Jordan, Sarah R. <1972> Learning to use contextual patterns in language processing Technical report no. 152, Computer Sciences Department, University of Wisconsin

Katz, J. <1966> The philosophy of language Harper and Row, New York

Kelley, K.L. <1967> Early syntactic acquisition Rand Corporation, Report no. P-3719

Lakoff, George <1971> "On generative semantics" in Steinberg, D.D., L.A. Jakobovits (eds.) Semantics, an interdisciplinary reader Cambridge University Press, pp.232-296

Lenneberg, E.H. <1962> "Understanding natural language without the ability to speak" Journal of abnormal and social psychology 65, pp.419-425

Limber, John <1973> "The genesis of complex sentences" in Moore, Timothy E. (ed.) Cognitive development and the acquisition of language The Academic Press, New York, pp.169-185

McConlogue, Karen, Robert F. Simmons <1965> "Analyzing English syntax using a pattern-learning parser" Communications of the Association for Computing Machinery 8:11, pp.687-698

McNeill, David <1971> "The capacity for the ontogenesis of grammar" in Slobin, Dan I. (ed.) The ontogenesis of grammar The Academic Press, New York, pp.17-40

McNeill, David, Nobuko B. McNeill <1968> "What does a child

mean when he says 'nc'?" In Zale, E.M. (ed.) Language and language behaviour Appleton-Century-Crofts, New York, pp.51-62

Moore, Timothy E. (ed.) <1973> Cognitive development and the acquisition of language The Academic Press, New York

Neisser, Ulric <1966> Cognitive psychology Appleton-Century-Crofts, New York

Olivier, Donald C. <1968> Stochastic grammars and language acquisition devices Ph.D. Dissertation, Harvard University

Reddy, D.R. <1971> "Speech recognition: Prospects for the seventies" Proceedings of IFIP 1, pp.12-18

Reddy, D.R., L.D. Erman, R.D. Fennel, and R.B.Neely <1973> "The Hearsay speech understanding system: An example of the recognition process" Proceedings of the third IJCAI pp.185-193

Schank, Roger C., Neil Goldman, Charles J. Reiger III, Chris Riesbeck <1973> "MARGIE: Memory, Analysis, Response Generation, and Inference in English" Proceedings of the IJCAI

Schlesinger, I.M. <1971> "Production of utterances and language acquisiticn" in Slobin, Dan I. (ed.) The ontogenesis cf grammar The Academic Press, New York, pp.63-101

Schwarcz, R.M. <1967> "Steps towards a model of linguistic performance: A preliminary sketch" Mechanical Translation 10, pp.39-52

Simmons, Robert, and J. Slocum <1972> "Generating English discourse from semantic networks" *Communications of the Association for Computing Machinery* 15:10 p.891

Slobel, Irwin <1973> "On Calibrating controlled cameras for perceiving 3-D scenes" *Proceedings of the third IJCAI*,, pp.648-657

Slobin, Dan I. (ed.) <1971a> *The ontogenesis of grammar* The Academic Press, New York

Slobin, Dan I. <1971b> "Developmental psycholinguistics" in Dingwall, William Orr (ed.) *A survey of linguistic science* William Orr Dingwall, Linguistics Program, University of Maryland, College Park, Maryland, pp.298-410

Staats, Arthur W. <1971> "Linguistic-mentalistic theory versus an explanatory S-R Learning theory of language development" in Slobin, Dan I. (ed.) *The ontogenesis of grammar* The Academic Press, New York, pp.103-150

Tappert, C.C., N.R. Dixon <1973> "A procedure for adaptive control of the interaction between acoustic classification and linguistic decoding in automatic recognition of continuous speech" *Proceedings of the third IJCAI* pp.173-184

Turing, A.M. <1950> "Computing machinery and intelligence" *Mind* 59, pp.433-460

Turner, Geoffrey G., and Bernard A. Mohan <1970> *A linguistic description and computer program for children's speech* Routledge and Kegan Paul, London

Wheatley, Jon <1970> Language and rules Mouton, The Hague

Wilks, Yorick <1973> Preference semantics Computer Science Department, School of Humanities and Sciences, Stanford University

Winograd, Terry <1971> Procedures as a representation for data in a computer Program for understanding natural language MIT Artificial Intelligence Laboratory, Cambridge, Massachussetts

Winograd, Terry <1972> Understanding natural language Academic Press, New York and London

Woods, W.A. <1969> Augmented transition networks for natural language analysis Computation Laboratory, Harvard University, Cambridge, Massachussetts

Appendix 1: The representation of VAS' Environment

```
((#AT #CENTRE (256 256 128)))
((#AT :HAND (0 0 256)))
((#IS :B1 #BLOCK))
((#IS :B2 #PYRAMID))
((#IS :B3 #BLOCK))
((#IS :B4 #PYRAMID))
((#IS :B5 #PYRAMID))
((#IS :B6 #BLOCK))
((#IS :B7 #BLOCK))
((#IS :B8 #BLOCK))
((#IS #RED #COLOR))
((#IS #BLUE #COLOR))
((#IS #GREEN #COLOR))
((#IS #WHITE #COLOR))
((#IS #BLACK #COLOR))
((#IS #RECTANGULAR #SHAPE))
((#IS #ROUND #SHAPE))
((#IS #POINTED #SHAPE))
((#IS :SHRDLU #ROBOT))
((#IS :FRIEND #PERSON))
((#IS :HAND #HAND))
((#AT :B1 (64 64 0)))
((#AT :B2 (64 64 64)))
((#AT :B3 (256 0 0)))
((#AT :B4 (416 416 1)))
```

```
((#AT :B5 (320 64 128)))
((#AT :B6 (0 192 0)))
((#AT :B7 (0 160 192)))
((#AT :B8 (192 416 0)))
((#SUPPORT :B1 :B2))
((#SUPPORT :B3 :B5))
((#SUPPORT :B6 :B7))
((#CLEARTOP :B2))
((#CLEARTOP :B4))
((#CLEARTOP :B5))
((#CLEARTOP :B7))
((#CLEARTOP :B8))
((#MANIP :B1))
((#MANIP :B2))
((#MANIP :B3))
((#MANIP :B4))
((#MANIP :B5))
((#MANIP :B6))
((#MANIP :B7))
((#MANIP :B8))
((#SUPPORT :TABLE :B1))
((#SUPPORT :TABLE :B3))
((#SUPPORT :BOX :B4))
((#SUPPORT :TABLE :B8))
((#SUPPORT :TABLE :B6))
((#SUPPORT :TABLE :BOX))
((#AT :BOX (384 384 0)))
```

```
((#IS :BOX #BOX))
((#IS :TABLE #TABLE))
((#CONTAIN :BOX :B4))
((#SHAPE :B1 #RECTANGULAR))
((#SHAPE :B3 #RECTANGULAR))
((#SHAPE :B2 #POINTED))
((#SHAPE :B4 #POINTED))
((#SHAPE :B5 #POINTED))
((#SHAPE :B6 #RECTANGULAR))
((#SHAPE :B7 #RECTANGULAR))
((#SHAPE :B8 #RECTANGULAR))
((#COLOR :B1 #RED))
((#COLOR :B2 #GREEN))
((#COLOR :B3 #GREEN))
((#COLOR :B4 #BLUE))
((#COLOR :B5 #RED))
((#COLOR :B6 #RED))
((#COLOR :B7 #GREEN))
((#COLOR :B8 #BLUE))
((#COLOR :BOX #WHITE))
((#COLOR :TABLE #BLACK))
((#CALL :SHRDLU SHRDLU))
((#CALL :FRIEND YOU))
```

Appendix 2: A fragment of corpus 1

1. I SEE BLOCKS AND LITTLE TRIANGLES - NO, THEY'RE NOT - RESTING ON A SQUARE, NO, IT'S AN OBLONG TABLE.
2. THE ROBOT'S HAND IS TO THE LEFT-HAND SIDE OF THE TABLE.
3. THERE IS A HOUSESHAPE MADE OF A RED CUBE, B3, TOPPED BY A GREEN PYRAMID, P1, DIRECTLY FACING ME TO THE LEFT ON THE TABLE.
4. BEHIND IT IS A CUBE, A RED CUBE, B2, MUCH LARGER THAN IT.
5. ABOVE IT, RESTING ON B2, IS A GREEN CUBE B1.
6. B1 IS PARTLY OFF B2, TOWARDS THE - FACING - TOWARDS ME.
7. BEHIND B1 AND B2, IN THE MIDDLE OF THE TABLE, IS A BLUE BLOCK, B5.
8. THIS BLOCK LOOKS LIKE A BLOCK.
9. IT IS STANDING ON END.
10. IN FRONT OF IT IS A GREEN CUBE B4.
11. IT APPEARS TO BE SQUARE.
12. ITS FRONT EDGE IS MATCHED WITH THE FRONT EDGE OF THE TABLE.
13. ON TOP OF IT, ON ITS RIGHT-HAND, BACK SIDE IS A LONG PYRAMID P2.
14. IT IS RED.
15. IT LOOKS LIKE A SET OF SKYSCRAPERS AND PRETEND CHURCHES.
16. THE ROBOT'S HAND IS DISTINCTLY UNHANDLIKE.
17. IT'S HANGING FROM A CROSS ABOVE THE TABLE, WELL ABOVE EVEN B1, PLUMB-LINE DOWN.

18. IT HAS A BAR ACROSS IT, ABOUT THE MIDDLE, NO, A BIT BELOW THE MIDDLE OF B1, HELD ON BY A LITTLE ROUND LOOP.
19. THE COLOURS IN THE WORLD ARE GREEN, RED, BLUE.
20. ONLY TWO OBJECTS HAVE SINGLE NAMES INSTEAD OF NUMBERS.
21. THEY ARE TABLE, AND BOX.
22. THERE IS A SPACE IN FRONT OF THE BOX WHERE YOU COULD PUT OTHER CUBES AND TRIANGLES IF ONE WISHED.
23. MOST OF THE REST OF THE TABLE IS FILLED UP WITH CUBES AND PYRAMIDS, ALTHOUGH VERY SMALL ONES COULD PROBABLY FIT INTO THE SPACES.
24. THE ONLY TWO BOXES - BLOCKS WHICH ARE NOT ALIGNED WITH SOMETHING ELSE, ARE B5 AND B3.
25. THEY ARE JUST SET DOWN - WELL, THEIR EDGES ARE PARALLEL TO THE EDGE OF THE TABLE IN EACH CASE, BUT NO EDGE IS EXACTLY ALIGNED WITH ANY OTHER EDGE.
26. IF I HAD TO LIVE IN A PLACE LIKE THAT I'D GO CRAZY.

Appendix 3: Foci corresponding to utterances in Appendix 2

1. (:B1 :B2 :B3 :B5 :B6 :B7 :B8 :BOX :TABLE :HAND)
2. (:HAND :B7)
3. (:B1 :B2 :B6 :TABLE)
4. (:B1 :B2 :B3 :B6 :B7)
5. (:B6 :B7 :HAND)
6. (:B6 :B7 :HAND)
7. (:B5 :B8 :BOX)
8. (:B5 :B8 :BOX)
9. (:B5 :B8 :BOX)
10. (:B3 :B5 :TABLE)
11. (:B3 :B5 :TABLE)
12. (:B3 :TABLE)
13. (:B3 :B5 :B8)
14. (:B3 :B5 :B8)
15. (:B1 :B2 :B3 :B5 :B6 :B7 :B8 :BOX :TABLE :HAND)
16. (:B6 :B7 :HAND)
17. NIL
18. (:B6 :B7 :HAND)
19. (:B1 :B2 :B3 :B5 :B6 :B7 :B8 :BOX :TABLE :HAND)
20. NIL
21. (:BOX :TABLE)
22. (:BOX :TABLE)
23. (:B1 :B2 :B3 :B5 :B6 :B7 :B8 :BOX :TABLE :HAND)
24. (:B5 :B8 :BOX :B1 :B2 :B6 :TABLE)

25. (:E5 :E8 :BOX :B1 :B2 :B6 :TABLE)

26. NIL

Appendix 4: Notation for describing ARTRANs

The notation described here is informal. The functions represented by the notation must be specified in detail in order to implement them with a specific semantic representation.

In figures 4.6 and 4.7, each node of the ARTRAN is represented by a box of the form $\to\boxed{n}\to$. Each arc is labelled with an input I, a process P, and a weight wi. The input is a string of characters, with blanks represented by "_". Processes may be:

"insert c in s" - Replace the slot described by s with concept c. The description of the slot may be arbitrarily complex.

"overlay p" - compare p with the Parse so far and replace any unfilled slot in the Parse with the corresponding concepts, if any, in p.

In this case, we have represented a slot by *n, where n is the relative left-to-right position of the slot in the Parse, since the structures built are simple predications.